Astronomical Image Processing. A1

Head of Experiment: Rupert Oulton

The following experiment guide is NOT intended to be a step-by-step manual for the experiment but rather it provides an overall introduction to the work 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 of course be cited in the bibliography of the report.

Computational Image Processing: A Deep Galaxy Survey

3rd Year Lab Module, Version 1.4: Revised September 2018, 2019.
Dr. D.L. Clements, N. Skrzypek, Y Unruh,
Astrophysics Group, Imperial College

1 Overview

The aim of this experiment is to use a deep optical CCD image of the extragalactic sky to determine the number counts of galaxies in the local universe (i.e., the 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 then 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. As with all computational physics, you should take care to develop well defined methods of testing your code and proving that it works as expected, before applying it to "new" data, for example by creating and analysing synthetic data containing well defined feature, or working with small sub-sections of an image that you can also characterise by hand.

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 (Figure 2) using the CCD mosaic camera and a "*Sloan r band filter*", which has a central wavelength at about 620nm (visible red). More information on the filter, camera and telescope can be found at :- https://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. While the observations with Spitzer are comparatively quite "old" at this time, no survey comparable to SWIRE has been done since then, because no space telescope matching the capabilities of Spitzer has been available.

Much work has already been done on the optical images obtained at the Kitt Peak telescope. Many separate exposures with different pointing have been combined and various instrumental effects and artefacts have been removed. The image you see in Figure 1 below is thus a fully "reduced" piece of data ready for scientific analysis, though there are still various features embedded in it, such as the bright star in the upper part, that are perhaps not ideal. Such data sets are constantly under review, however the expense in terms of observation and people time does mean that often an "older" piece of data may still be the best available. This is the case here since no deeper optical data for this region of the SWIRE survey has been taken. We actively review the data used in this experiment (last review September 2018) to make sure that it remains topical.

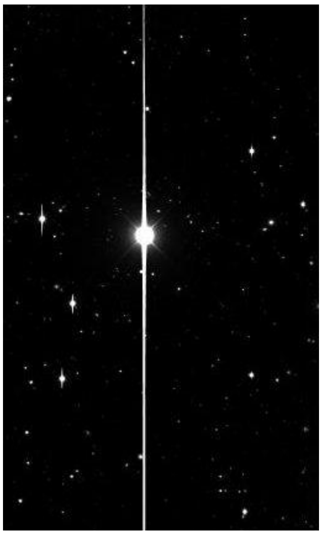

Figure 1. The optical survey field used in this experiment. Note the many objects, almost all of which are galaxies, the bright star in the upper centre with CCD bleeding across nearby pixels and associated column of "noise" pixels, and the variability of noise with position.

Figure 2. The 4m telescope at Kitt Peak Observatory

2.1. Apparent Magnitudes

The brightness of astronomical objects as observed from Earth is typically measured in what are called "*apparent magnitudes*" (see e.g. Walker, 1989, p. 12). This system derives from observations made by early Greek astronomers who classified stars into several different magnitude classes. Since the eye's reaction to brightness is logarithmic (elegantly compressing dynamic range, and allowing you to process images in both bright sunlight and near darkness), the magnitude system is itself logarithmic. The system was formalised by Norman Pogson in the 1850s. The standard definition relating the difference in magnitude between two objects to the flux (a measure of energy per time flowing through an area with SI units of W m^{-2}) received from them is given by:

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

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 for astronomical magnitudes:

- an object 2.5 magnitudes brighter than another emits 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 and wavelength ranges.

The standard definition of the magnitude is (somewhat counter intuitively) entirely relative: it merely addresses the magnitude difference between two objects. To allow sources observed at different telescopes to be directly compared, we must go further than Pogson's definition given in Eq. 1. Firstly, the wavelengths of light received by the telescope need to be standardized. This need to standardise wavelength ranges leads to the use of specific sets of filters in the optical and near-infrared part of the spectrum. These include the Johnson filters (e.g. Walker, 1989; Bessel, 1990) – U B V R I - and the more modern SDSS filters (see the Sloan Digital Sky Survey website for more details) - *u g r i z* - one of which is used in this investigation. In addition to the filter set used, we must also choose a "standard" 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 with whatever filter is being used to limit the observation range. There is also the AB magnitude system, where a specific flux density, 3631 × 10^{-26} W m^{-2} Hz^{-1} is defined to be zero magnitude. In the sometimes-strange systems of units used by astronomers, this is 3631 Janskys, since 1 Jy is defined to be 10^{-26} W m^{-2} Hz^{-1}. As an aside, "older" scientific disciplines like astronomy have often accumulated unit oddities like this as they pre-date the establishment of SI and other more "decade based" scales of measurement. The Jy in fact is still used in astronomy since astronomers don't like having to use very small numbers all the time. The same is also true of the magnitude scale.

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

3. CCD Astronomy

Most optical astronomy today relies on the use of electronic sensors such as Charge Coupled Devices (e.g. Walker, 1989) - so called CCDs - rather than photographic plates. These devices have considerable advantages over traditional "wet" photographic techniques, but there are also some limitations. 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 individual photons falling on it than a photographic plate, or the eye. A typical CCD can detect as many as 90% of the photons incident on it, while a photographic plate is at best, 10% sensitive. Thus, by switching a telescope from a photographic detector to a CCD, you are effectively multiplying its photon collecting capabilities by nearly an order of magnitude.

As an interesting aside, consider the implication of the move to CCD detectors. 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 photon counting detectors that are linear over a large range.**

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 from an object, i.e., 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 the source brightness must be very carefully calibrated. However, CCDs are not perfectly linear. 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 16-bit counter, and so for counts-per-pixel approaching the maximum allowed of 2^{16} = 65,536, the linearity breaks down. The practical aspect of this is that sources with recorded signal levels greater than ~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 start to "bloom" - with the electrons it has collected effectively overflowing and spreading to nearby pixels. These can then overfill themselves and contaminate still more pixels. This overflowing takes place preferentially in specific directions on the chip (generally along the rows and columns of the CCD - where data is sequentially read out), leading to the sort of unphysical and unwanted line seen running from top to bottom in Figure 1, associated with the very 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 easily by a computer, and stored directly to tape or disk at the telescope and then copied and distributed across the world with no loss of fidelity. It is quite common for astronomers to book telescope time and run the entire observation and data collection process remotely as a result. With photographic plates there were many additional stages required, including developing and copying of plates, and the original media itself needed to be treated very carefully. Usually, copies rather than original plates would have been used and they would have been susceptible to ageing and a loss of resolution in each copying process. Here, you will be using CCD data in exactly the same form as used by professional astronomers. There are however various additional processing stages required before an image like the one shown in Fig. 1 is obtained.

These processing steps include calibration, and the combination of many different sub-exposures into the one image. This processing leads to a number of effects, most obviously the increased noise towards the edge of the main image you will analyse, since this part of the image includes fewer sub-exposures and thus has higher statistical noise, and the jagged corners, where the edges of the various sub-exposures fail to overlap (most easily seen in the top right and bottom left of the image).

You may wish to think about what instrumental effects need to be corrected for before a final image such as this is produced, and discuss with the demonstrator why numerous sub-exposures might be needed.

4. Galaxy Surveys

Galaxy surveys (see, e.g., 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 - i.e. exposures 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 some "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. This is a technique known as "*number counts*" and was first used by William Herschel in 1784 (Herschel, 1784) to determine the disk-like structure of our own galaxy through counting stars (by eye). The technique is still used today in extragalactic astronomy at all wavelengths, from X-ray to radio, and is usually the first thing that extragalactic astronomers do as soon as a new wavelength range or sensitivity regime is made available to them by a new observatory (see e.g. Clements et al., 2010) for work done in the Blackett Lab.

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:

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

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

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

Equation 2 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 wider universe. Indeed, since galaxies are distributed much more uniformly (on the large scale) 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 e.g. 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 - and 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 e.g. 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 always "incomplete" to some degree. The level of incompleteness is eventually 100% at the detection limit of the survey. In principle, 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, for two quite distinct reasons. Firstly, bright sources are rare, and thus you will have relatively few of them in your catalogue. Simple Poisson statistics applies to counting these galaxies, and so the error on the number of counts is simply $\sqrt{(N)}$. Consideration of this error will allow you to account for the statistical effects of the rare bright

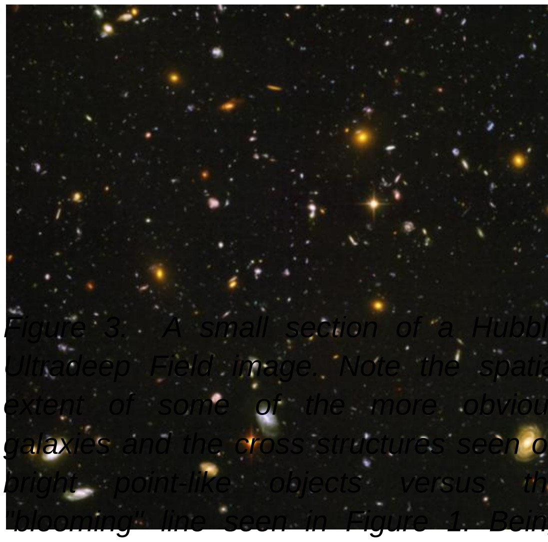

Figure 3. A small section of a Hubble Ultradeep Field image. Note the spatial extent of some of the more obvious galaxies and the cross structures seen on bright point-like objects versus the "blooming" line seen in Figure 1. Being space based, Hubble is far less limited by blurring caused by an intervening medium, but other effects such as diffraction then become more obvious.

sources. However, a second factor also 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 (not purely point-like) in the images. Figure 3 shows some spectacular examples from The Hubble Ultradeep Field used for, among other things, faint number counts of galaxies.

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 if it lies outside the aperture. This will make the brighter objects appear to be fainter in the catalogue 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 our "local" galactic stars. Bright ones can be clearly picked out (e.g. in figure 3), 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 is time to discuss the experiment itself, and how you'll produce galaxy catalogues from the data provided.

5 The Experiment

The goal of this experiment is to take a large, deep optical image of the sky, and develop a computer code (or multiple pieces of code) that will produce a catalogue of all the objects in this image, together with their photometry (optical flux), allowing studies of the galaxy population to be undertaken. To achieve this, software that performs astronomical image processing must be written, tested and applied by you.

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

1. Read the astronomical image.

You need to load an image into the computer in a form that allows further processing steps to 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 will measure the flux 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 catalogue of the objects you have detected and the amount of light emitted by them.

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

5.1 Acquiring the data

The data for this experiment (~22.6 Meg) 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 (a block of non-image information at the start of the file) which describes the data, how it was obtained, and, in some cases, how it was processed. This so-called "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 file into memory so that your programmes can be applied to the data. There are several libraries of pre-built and well tested 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 and desire (or otherwise) to learn new programming languages.

MatLab is a high-level data manipulation package widely used in research environments. It comes with a broad suite of complex routines built in, but also has something of a "learning curve". It is available on the machines in 3rd year lab and is a popular and well-supported system. You can also use other programming languages such as **C, C++, Python**, etc. You might choose to work with MatLab to learn a new "technical" skill, or with Python if you think that will be "faster", the choice is yours. Code developed in C or C++ can run considerably faster than MAtlab or Python due to the way it is pre-compiled, however, you should probably avoid C and C++ unless you are already quite familiar and fluent with them.

5.2.1 In Matlab

College has site licences for MatLab which may allow you to install a copy on your own laptop. You can find out more about this at the following URL.

https://uk.mathworks.com/academia/tah-portal/imperial-college-london-600177.html

Once you have access to MatLab (on your own or a College machine) 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.

5.2.2 In C++

The most commonly used libraries for handling FITS files are in ***CFITSIO***, which is written in C but callable by C++ and other languages, and ***CCFITS*** (currently at version 2.5 as of September 2018), which is a native C++ interface to the CFITSIO library. This software can be downloaded from :-

https://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.

5.2.3 In Python

We recommend using the ***astropy*** package to access FITS files with Python. The package is loaded in a standard way:

>>>from astropy.io import fits

This will enable you to open an existing FITS file. For the purposes of this experiment, you will need very little ***astropy*** functionality beyond opening the file. Otherwise, standard ***numpy*** operations are sufficient.

You can check the contents of the FITS file in the following way:

```
>>>hdulist = fits.open("/path/to/mosaic.fits")
Filename: mosaic.fits
No.   Name      Type          Cards   Dimensions     Format
0     PRIMARY   PrimaryHDU    179     (2570, 4611)   int16
```

This is useful if, for example, your FITS file contains multiple data tables or additional noise or S/N maps (not relevant for this experiment as there is only one component - a.k.a. Header Data Unit, or HDU - in the FITS file). If you use *fits.open()*, you should always close the file when you have finished working with it using *hdulist.close()*.

The FITS header can be obtained using:

```
>>>hdulist[0].header
```

or, directly from the filepath

```
>>>fits.getheader("/path/to/mosaic.fits",0)
```

and values can be accessed either using indexing (list) or keywords (dictionary). The file header can also be viewed using DS9 or equivalent software.

The FITS data can be accessed using

```
>>>hdulist[0].data
```

or, directly from the filepath with the equivalent *fits.getdata* command. These commands return a numpy ndarray object, which contains the count value at every pixel in the image.

For reference, complete, thorough astropy documentation is available at:

http://docs.astropy.org/en/stable/io/fits/.

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 might find useful, especially if working in C or C++. You can view FITS files on Windows using software such as FITSVIEW (https://www.nrao.edu/software/fitsview/) or the more sophisticated ds9 (http://ds9.si.edu/site/Home.html) which is also available for Linux and MacOS (most real astronomers use ds9, whose name derives from a popular television science fiction series). 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 (e.g., using CFITSIO) and then read it into the visualisation programme.

5.2.5 A Tour of the Image

Once you have ds9 or another FITS viewer working you should have a look at the image by eye. This will help familiarise yourself with the data and will also demonstrate that there are some things that the eye can see that an automated system might find much harder to identify (and vice versa of course). To be able to see these features you will need to set the scaling of the image viewer to an optimum value. For ds9 you should set 'scale' to 'zscale', which allows faint sources to be seen in the presence of much brighter ones, and press the 'log' button. This sets the scaling so you are looking at the log of the flux value, and allows a wide dynamic range of features to be seen.

Once you do this, the most obvious feature you see is the bright star in the upper centre of the image. This is sufficiently bright that it has 'bled out' into the surrounding CCD area and into the bright line running from top to bottom of the image. There are also horizontal bleeds at the bottom of the image corresponding to the edge of the CCD frame in the various different sub-exposures that were combined to make this final image. There are also a number of other bright stars with some level of bleeding. These can be seen if you change the scaling to just select bright sources (in ds9 you can do this by setting scaling to 'min max').

Objects that show any sign of bleeding, which can easily be spotted by eye but which are less easy to find automatically, have gone very non-linear and do not provide anything useful for this experiment. You should find methods of excluding them from the analysis.

Next, go back to normal 'zscale' scaling and look not at the objects, but at the noise in the blank sky between them – you may want to zoom in to get a closer look. What you find is not blank black sky, but a general fuzziness with a value significantly above zero. The non-zero value is a result of sky background – ie. the fact that the sky is not in fact ever perfectly dark at night. The fuzziness you see here also has a contribution from noise in the detector. These are both things that you will have to deal with in the later processing. As noted above, you should also be able to see that the noise is greater towards the edges of the image. This is because these parts of the image were not observed by as many sub-exposures.

Looking back at the full image you will also be able to see that the sky background appears lowest in the bottom right of the image. Such gradients in the background are fairly common and are often due to the detailed scattering of light from a bright object in the field (such as the bright star), or somewhere else in the sky at the time of the observation (the Moon is one such source).

In general, the sky background is not a fixed value in any image. This is something that you may wish to develop strategies to both investigate and account for in your analysis.

5.2.6 The Statistics of the Image

Once you have read the data into your chosen analysis tool, one of the first things you should do is look at a simple histogram of the pixel values. This will give you a general idea of the statistics of the image. What you should see is a Gaussian distribution of points around some mean value with a tail of histogram bins extending beyond the Gaussian to more positive values. You may need to suppress very high pixel values in the histogram, due to the bright stars and blooming, to be able to see this.

The peak of the Gaussian represents the average background counts in the image. The spread of this Gaussian represents the noise in the image. The tail of bins extending to larger positive values are the pixels that contain photons from astronomical sources, and are thus the pixels you are particularly interested in for detecting "real" objects.

Looking at the distribution of pixel values, and considering how many pixels there are in the image, think about how many standard deviations away from the mean background value you have to be to be sure that a pixel has a high value because of a source, and not just because of the noise. Discuss this with your demonstrator.

5.3 Source Detection

The first step towards producing a galaxy catalogue 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 - provided it is not an extended object spread over many pixels. The photometry of this source can be extracted and saved to your catalogue. You then need to indicate that this bright source has been dealt with so that you do not "rediscover" it and 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 (e.g. each source as it has been added to the catalogue) or areas which are deemed "bad" such as those associated with a bright star or blooming. This could be a very simple data structure, for example a "logical" array of 1 / 0 or True / False values.

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 a very important step in your analysis 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. This annulus may contain real sources, especially if the annulus chosen is large, which must be excluded to make an accurate determination of the background.

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 catalogue and number counts.

What problems would you expect to 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 your source to source magnitude. To do this you first convert your counts into instrumental magnitudes:

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

Instrumental magnitudes are then converted to calibrated magnitudes by adding the instrumental zero point ZP_{inst}:

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

or equivalently:

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

5.6. Producing the Catalogue.

The results that you calculate for each object detected need to be stored in a catalogue. This can be a simple ASCII file that can be read into another programme for further 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 this data you may 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 (i.e., $N(< m)$ 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 Implementation Issues

6.1 Developing, testing, documenting and archiving your code

To make sure that your code is doing what you intend, you shouldn't immediately apply it to the astronomical data even though that may be tempting. Instead, first test your code by applying it to well understood "fake" or synthetic data where you know what the result should be. Start with the most obvious limiting cases, for example, make sure your code does not find objects in a fake image that is all zeros, all the same value, or just random noise. Then add a simple image of a source – best simulated using a Gaussian "blob" with a FWHM size of ~1arcsecond, matching the best performance from a ground based telescope without corrective optics*.

When you are sufficiently confident that your code is doing what it is meant to do, (and have the results needed to demonstrate this in your report) you can begin to test it on the real data and get rapid feedback on the results by working with small subsections of the larger image. A run on the whole image can take up to an hour for some code, so testing it on a subsection of the image can greatly speed up the testing and debugging part of the process.

Once your code is running to your satisfaction you may wish to think about how to determine how well it is recovering sources of a given brightness form the image. One way to do this is to add simulated sources of a known flux to the image, or subsections of the image, and then to see if your code successfully detects them and accurately recovers their flux. In this way you can actually determine the incompleteness of your catalogue as a function of magnitude.

In developing your code, it is a good idea to break the larger problem down into a set of simple subtasks and then to develop these individually. The block diagram for how these subtasks fit together is also extremely useful for describing your code and should feature in the final lab report. As you go through the development of your code you will need to make sure you keep track of different versions of both code and data files, and that your code is commented so that you and anybody else (particularly a demonstrator or marker) can understand what it is doing. Details of these different versions should be entered into your lab book, along with details of any tests of your code.

Output files for source catalogues etc. should have sensible names that allow you to understand what is in the file. Catalog1.dat and Catalog2.dat are not helpful names. You may also want to use a code archiving site such as GitHub to coordinate code development and provide access for marking. You should provide a link to any such repository in your report so that your assessor can easily look at it and potentially run it. You will also need to include comments in the code detailing any dependencies (e.g. Astropy) as well as including essentials such as date, author, and functionality.

*Footnote – While you might think that the diffraction-limited resolution of a 4m telescope would be given by the Airy disk – $1.22\lambda/D$ – it is in fact turbulence in the atmosphere that is the key limiting factor. For most large field instruments, such as the one used here, 1 arcsecond resolution is the best angular resolution that can easily be achieved without using "active" correction.

6.2 Astronomical Issues

There are several niceties of the astronomical data that will have to be considered in your analysis which go beyond the somewhat idealised discussion above. Firstly, galaxies are not simply unresolved point sources. They are extended objects that in many cases will be resolved by these observations. Their shapes will not be circularly symmetric, so they may appear as extended elliptical shapes and in some cases substructure within them can be discernible in the image, such as a bright core with extended disk.

How will your extraction routine cope with this? Can you test your extraction routine on simulated galaxy shapes to show that it can cope with them?

Secondly, in some cases galaxies will be quite close together. This may be the result of chance alignments along the line of sight, but they may also be genuine physical associations that can lead to galaxy interactions and mergers. How will your extraction routine cope with close pairs of galaxies? Can you simulate such instances and demonstrate that your code can cope with them? You can use ds9 or any other FITS viewing tool to look at the image and see how common well resolved galaxies and galaxy pairs are in the image. On this basis you can assess how much of a problem these issues might be if left uncorrected.

7. 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 – still today (2018) one of the best references for optical galaxy counts despite its age) 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? What other possible explanations might there be? Discuss your conclusions in your report.

Report File Size Problems.

If you include a copy of the main image in your report, please make sure that this is compressed or resized in some way to avoid problems with the Blackboard, Sharepoint or Email systems. You should aim for a final PDF report no larger than 10Mb.

8. References and Useful Documents.

Bessel, M.S., 1990, Publications of the Astronomy Society of the Pacific, 102, 1181.

Clements, D.L., et al., 2010, A&A, 518, 8.

Clements, D.L., 2014, Infrared Astronomy: Seeing the Heat, pub. CRC Press

Herschel, W., 1784, Royal Society Philosophical Transactions, 74, 437

Lonsdale, C.J. (SWIRE survey), 2003, Publications of the Astronomy Society of the Pacific, 115, 897.

Narlikar, J.V., 1983, Introduction to Cosmology, pub. Jones & Bartlett.

Peacock, J.A., Cosmological Physics, 1999, pub. Cambridge University Press.

Walker, G., 1989, Astronomical Observations, pub. Cambridge University Press.

Yasuda et al., 2001, Galaxy Number Counts from the Sloan Digital Sky Survey Commissioning Data, The Astronomical Journal, vol. 122, pp 1104 - 1124.

Footnote, for the sharp eyed, in the Herschel reference above, the "1784" does not refer to a page number. It is for reasons like this that Astrophysics has acquired its nomenclature over a long period of time.

Manuals and Observatory information.

Kitt Peak Mosaic Camera Manual : https://www.noao.edu/kpno/manuals/dim/

Sloan Digital Sky Survey : https://www.sdss.org/

Spitzer Space Telescope Website : http://www.spitzer.caltech.edu/

CFITSIO: https://heasarc.gsfc.nasa.gov/docs/software/fitsio/fitsio.html

CCFITS: https://heasarc.gsfc.nasa.gov/docs/software/fitsio/ccfits/

FITSVIEW: https://www.nrao.edu/software/fitsview/

SAO/ds9: http://ds9.si.edu/site/Home.html

Literature Search Tool

ADS: http://adsabs.harvard.edu/abstract_service.html