Determining the Number Counts of Galaxies in the Local Universe Using Computational Image Processing

Isabel Jones

***Abstract*—In this experiment we used computational image processing to study an image of the extragalactic sky taken by a CCD mosaic camera at the Kitt Peak telescope. We designed, tested, and used an algorithm that could detect objects in the sky and calculate their magnitudes. Using this we produced a number count plot i.e. the number of galaxies less than a given magnitude against that magnitude. We compared this to the theoretical result that were derived using the assumption of a Euclidian universe. The expected gradient from the theoretical prediction was 0.6. We achieved a gradient of 0.559 which was a 0.07% difference to the theoretical prediction. We discuss the objects that don't fall into this predicted trend and compare it to the work of Yasuda *et al.* The graph follows the theoretical trend for magnitudes between 16 and 21. We finish by discussing ways our algorithm could be improved.**

I. Introduction

Observing astronomical objects by grouping sources into their apparent magnitudes has been done since the early Greek astronomers when looking at stars. Society has since developed better technologies and cameras that can capture more information from stars/ galaxies. Charge-coupled devices, namely, are one of these technologies and are invaluable in digital imaging. They have high quantum efficiencies and so are used for most applications by astronomers between infrared and UV. CCD's work by representing pixels with capacitors. The array of capacitors allow a transfer of electric charge between them, which can be read [1]. This device proves much more efficient than traditional techniques as it can detect up to 90% of photons on the plate.

In this experiment, we looked at an image of the extragalactic sky that was taken using a CCD mosaic camera at the Kitt Peak 4m telescope. The "Sloan r band filter" was used to standardize the range of the wavelength. This filter has a central wavelength in the visible red region of 620nm. We then aimed to measure the magnitudes of the brightness of the sources in this image, to eventually generate a plot for the number of sources brighter than a given magnitude.

II. Theory

A. Number counts

The difference in apparent magnitudes of two sources m_1 and m_2 can be defined as,

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

Where f_1 and f_2 are the fluxes of the two sources given by the energy flowing through the area per unit time. The negative sign in Equation 1 tells us that a galaxy with a larger brightness will actually have a smaller apparent magnitude than a galaxy with a smaller brightness.

To be able to compare sources detected from different telescopes, we will need an equation that can produce the magnitude of one object and not the difference between two like in Equation 1. One method to do this was introduced by the Greek astronomer, Hipparchus. He achieved the results by first defining the magnitude of the star Vega to be zero. This definition allows the magnitude to be written as,

$$m = -2.5log_{10}(f) + c \tag{2}$$

where the constant c depends on the magnitude of Vega (it is a 'zero point' constant). The inverse square relation of the flux can be used in the absence of absorption,

$$f = \frac{L}{4\pi r^2} \tag{3}$$

where L is the luminosity of an object observed at a distance r. We can then use the expression for the number of galaxies within in a radius r and with a number density ρ,

$$N(r) = \frac{4}{3}\pi r^2 \rho \tag{4}$$

Equation 4 is based on the assumption that the universe is Euclidian, and therefore homogeneous and isotropic. Thus, the equation we are deriving assumes the galaxies to be uniformly distributed.

By subbing Equation 3 into Equation 2, we can rearrange for an expression of r and then substitute this into Equation 4 to get,

$$\mathrm{N(m)} = \mathrm{c} \times 10^{0.6m} \tag{5}$$

Here c is a constant.

Then we simply take the logarithm of both sides to get the desired expression of the number of galaxies below a certain magnitude m,

$$log_{10}(\mathrm{N(m)}) = 0.6\mathrm{m} + \mathrm{constant} \tag{6}$$

This is the relationship we worked towards achieving in this experiment [2].

B. Background noise and bleeding

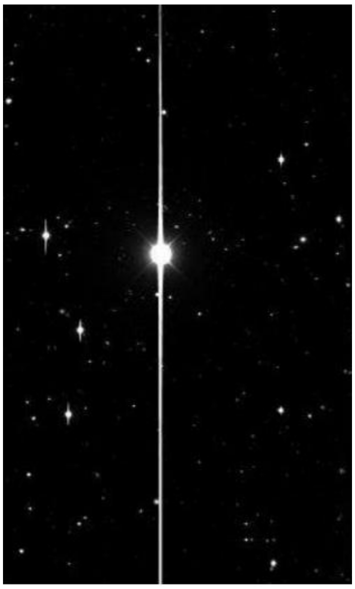

Figure 1 shows the original image of the extragalactic sky taken with the CCD mosaic camera at the Kitt Peak telescope. The image shows signs of bleeding in the brightest objects.

Although CCDs provide a great deal of accuracy and are often superior to their rival technologies, they will still produce a few errors in the image.

First, CCDs are only linear up to a point, so they are capped at the number of electrons they can collect. Since the number of electrons corresponds to the pixel brightness, the brightness's detected will also be limited. Therefore, there will be unreliable flux values for sources above fifty thousand counts per pixel.

Moreover, when a pixel becomes too full it will start to bleed or "bloom". This occurs when electrons spill over to neighboring plates as the plate cannot store them above the cut-off number, and so neighboring pixels become brighter. This leads to the vertical line shown in Figure 1.

Another feature we want to remove from our image is the noise. In processing the image, sub-exposures are used to try and reduce this noise. However, this processing fails to reduce the defects along the borders of the image as these regions have fewer sup-exposures [3].

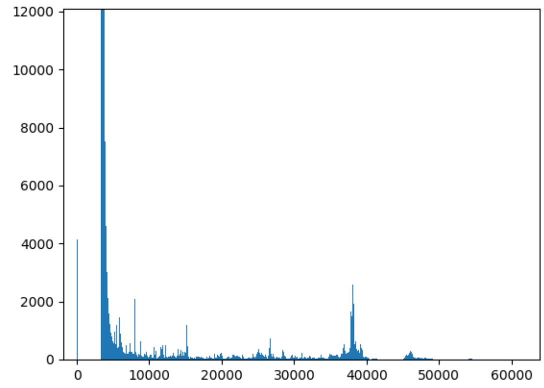

Figure 2 shows a histogram of the pixel values. The first large spike is due to the noise in the image and the second largest spike occurs due to bleeding.

Using Figure 2, we sliced our data to look at the first peak which represents the noise, as this is something we would later want to eradicate. We plotted the noise to show it follows a Gaussian distribution as shown in Figure 3. From this we could use the curve-fit function in python to find values for the mean and standard deviation of the noise. We found the mean pixel value of the noise to equal 3418.7 and the standard deviation was 17.3.

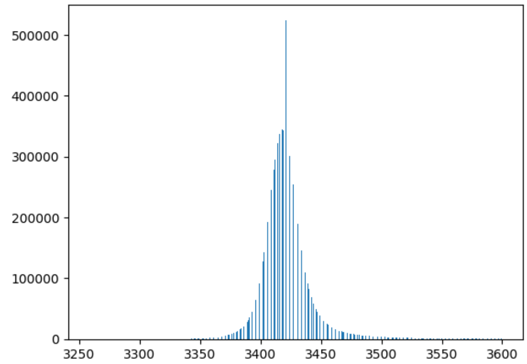

Figure 3 shows that a histogram plot of the noise follows a Gaussian distribution.

III. Experimental Method

First, we displayed the image on a log scale in Fitsview. This allowed us to view the many galaxies present and scale the image as our eyes would if looking at the sky. Then, to see the effects of the noise and bleeding we plotted the pixel count on a histogram as shown in Figure 2.

Next, we masked the image to remove unwanted bleeding effects and the defects around the borders. This was achieved using a photoshop software called GIMP. Using this we produced a masked layer whereby the brightest galaxies that exhibited blooming and the borders of the image were removed. This layer was then multiplied this by the original image in python to get the correct scaling.

After masking the unwanted part of the image, we wanted to find the brightest pixels as we assumed these would be the centre of the galaxies we wanted to measure. To do this we implemented the code as shown in Figure 4.

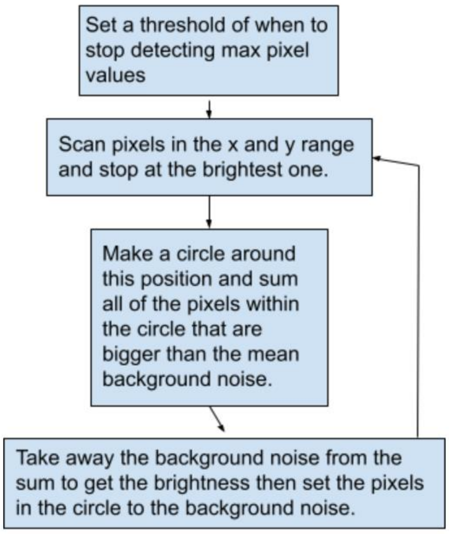

Figure 4 is a flow diagram of the computational process of finding sources and calculating their brightness. The mean used was determined from the gaussian distribution.

To choose the threshold value used in Figure 4, we had to find the balance between choosing a value that wasn't too high, which would miss sources below a certain value, whilst also not counting background defects as galaxies by choosing a value too low. After testing without a threshold and detecting a lot of 'galaxies' that in reality were part the noise, we tested on a pixel value of 3540 to be the threshold. Figure 5 shows a small section of the astronomical image that we used to test how the algorithm worked with a circle of radius 6 pixels. The image shows that the radius does not include the whole galaxy for larger galaxies, thus impacting the brightness. Therefore, to produce our final graph we chose a radius of 10 pixels as above this would be too large an area for our program to be able to compute in the time frame. Moreover, since we subtracted the mean background value from each pixel value before we summed them for the total brightness, a large radius that includes background noise should not affect the final magnitude of the object.

The image also shows that some galaxies have not been picked up by our algorithm implying we had to lower the threshold. We thus chose a final threshold to be the mean of the noise plus three standard deviations, as any lower became too difficult for our computers to run due to the large amount of data.

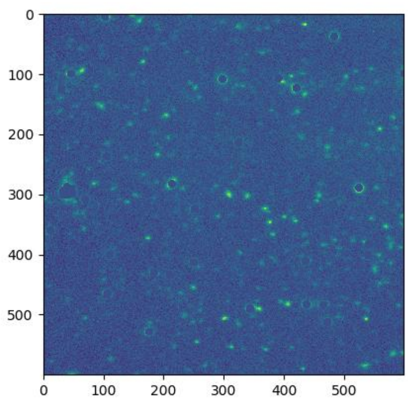

Figure 5 shows a section of the studied image that the code has been applied to. The radius of the circles was 6 pixels and the threshold value of the pixels detected was 3540. The galaxies collected can be seen by the dark circles within the light patches.

Once we calculated the brightness of each galaxy, we found the magnitude of a galaxy, m, using the instrumental zero point ZP_{inst}, which is stored in the Fits file. Then by using the equation,

$$m = ZP_{inst} - 2.5log_{10}(counts) \tag{7}$$

where the counts are the brightness i.e. the sum of the pixel values within the circle after subtracting the mean background noise.

Then, we could put into our code a line to find the number of galaxies with a magnitude less than a given magnitude and plot the logarithm of this number count against the magnitudes. The errors of the number count, N, were $\sqrt{N}$. We could then use the error propagation equation to calculate the errors of the logarithm of N and to plot the error bars on the graph.

We used a curve-fit function to compare our results to the theoretical results in Equation 6.

IV. Results, Errors, and Discussion

Using the experimental method, we plotted the graph shown in Figure 6. This produced a gradient of 0.559 which only differs by 0.07% to the expected value of 0.6 from Equation 6. We produced a constant of value 6.57. Our algorithm generated around fourteen thousand objects.

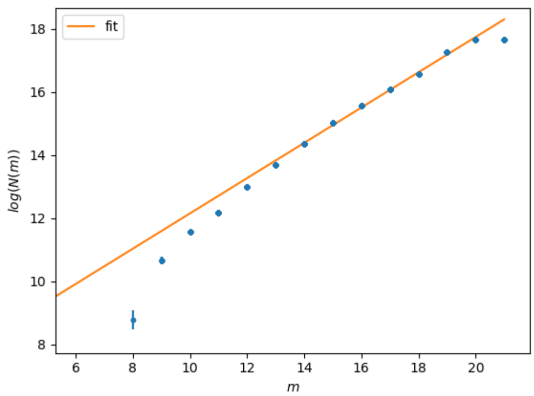

Figure 6 shows a plot of log(N(m)) against the magnitudes. This is also known as the Number Count plot.

Though our results are very close to the theoretical expectation, we also have a few galaxies outside of the line not included by their error bars. These values of lower magnitudes are the brightest objects. These are likely to be stars at close proximity to the earth. The fact there is not perfect alignment to the theoretical condition for these objects could be due to the assumption we made of the Euclidean universe breaking down for these brighter/ closer objects. Moreover, the galaxies observed very far away are much older than the close galaxies. This means the properties of the galaxies could have changed over time, explaining the deviation from the trend for the largest magnitude as well as the smallest.

Yasuda *et al.* have results that say the magnitudes between 16 and 21 follow the theory the most accurately [4], which is inline with our results in Figure 6.

V. Conclusion

In this experiment, we aimed to produce a number count plot for the galaxies in an image of the local universe captured with a CCD mosaic camera at the Kitt Peak telescope. We managed to achieve results that were very close to the theoretical predictions and to the thorough research carried out by Yasuda *et al.*

To improve our experiment, we could improve the algorithm used to detect stars. If a star contained multiple bright pixels of equal magnitude, several circles could be drawn for the same star. Therefore, our algorithm could effectively count one object more than once. Moreover, our algorithm could account for the different shapes of the objects as some are elliptical. Finally, our algorithm took a very long time to scan through image, implying it was not very efficient. With more time, these things could be developed.

Overall, the experiment was a success as we produced the desired results.

References

[1] "5: MOS Capacitor and MOSFET - Semiconductor Devices: Physics and Technology, 3rd Edition [Book]." https://www.oreilly.com/library/view/semiconductor-devices-physics/9780470537947/13_chap05.html (accessed Mar. 07, 2021).

[2] A. Campos, "Galaxy Number Counts," 1995.

[3] "IEEE Xplore Full-Text PDF:" https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7919164 (accessed Mar. 07, 2021).

[4] N. Yasuda *et al.*, "Galaxy Number Counts from the Sloan Digital Sky Survey Commissioning Data 1," 2001.

Appendix

Code used for analysis can be found at: https://github.com/iaj18/Astronomy-labs/blob/main/Code