Computational Image Processing for a Deep Galaxy Survey

James Crawford

***Abstract*—The galaxy count and information about the distribution of the galaxies sizes and brightness of some given image was determined by means of an imaging processing program. The image was taken using a CCD from the Spitzer Space Telescope. This program worked by iterating over the brightest pixels that qualified as a galaxy pixel. For each pixel, an expanding circle was used to estimate the size and brightness of each galaxy, and a list of each galaxy and information about its brightness, size and location could be returned. The count count magnitude plot could be determined, and the value of 0.42 ± 0.01 was obtained and 1127 galaxies were found.**

I. Introduction

In the last hundred years, our understanding of the universe has changed dramatically, partly because of the recent development of more sophisticated equipment and telescopes that have allowed us to probe deeper into the universe. In particular it became established that the Milky Way is just one of thousands of millions of other galaxies that exist in clusters in the universe [1]. Because of this, it has become a necessity to have programs that are able to perform galaxy counts of optical images, and extract information from these images that might be used to study the nature of the evolution of our universe.

The program in this experiment does exactly that: given an optical image taken with the Kitt Peak 4 m telescope, it detects the number of galaxies in the image, and records information about the brightness, position and composition of each one [2]. This was done by considering all the pixels that exceed some local threshold, and estimating each galaxy as a circular image. From the collected pixels, information about each galaxy could be determined.

By performing a magnitude count plot, the results could be compared to recent studies to see how effective the program was in correctly identifying galaxies.

II. Theory

Within our universe there are thousands of millions of galaxies [1]. They vary in in shape and size and distribution but can be largely grouped into elliptical, lenticular, spiral and irregular. Spiral as the name suggests are galaxies that are large discs with spiral arms and a central bulge. Lenticular are similar but without the spiral arms. Irregular are galaxies that do not fit into the other categories and will not dealt with in this paper. Most importantly are the elliptical galaxies which will dealt with in this paper [1].

They are elliptical in shape and have a light output that is concentrated in the centre, and decreases constantly with increasing radii. Fig. 1 shows an example of an elliptical galaxy.

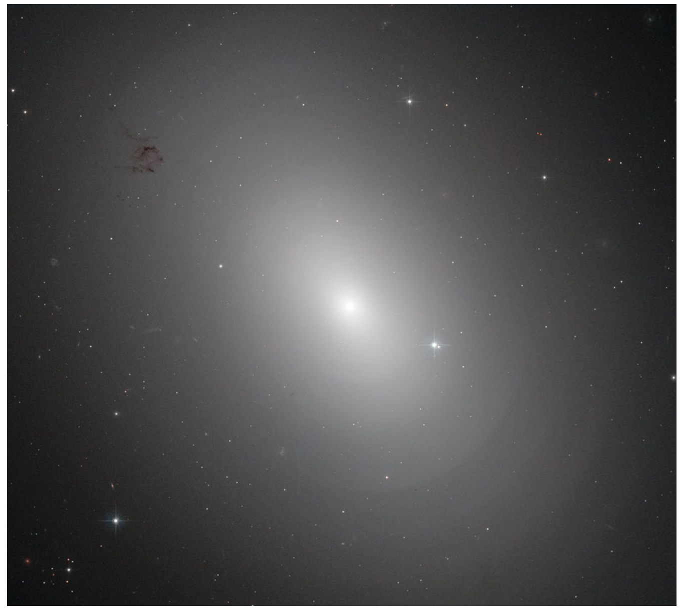

Fig. 1. Example of an elliptical galaxy [3].

Suppose that these galaxies and distributed uniformly in Euclidean space. The number $N(>S)$ of galaxies greater than some flux density S in some solid angle Ω is given by

$$N(>S)dL = \frac{\Omega}{3} r^3 N dL \tag{1}$$

where L is some luminosity function, r is some limiting distance and the inverse square law is used [4]. By integrating (1) over the lumonisity function and by using the following definition of magnitude:

$$m = c - 2.5 \log S, \tag{2}$$

an expression for $\log N(>m)$ can be obtained as the following:

$$\log N(>m) = 0.6m + c, \tag{3}$$

which is the expected distribution of the galaxies [4]. While photographic plates used to be used in telescopes, CCD cameras are now used to take images. They provide a high quantum efficiency, and below some saturation level provide a linear response in that each photon will produce an electron [2]. After an exposure, charge travels from cell to cell before being recorded by some computer, which leads to bleeding since charge overflows up or a down cells [5]. An example of a star bleeding is shown in Fig. 2.

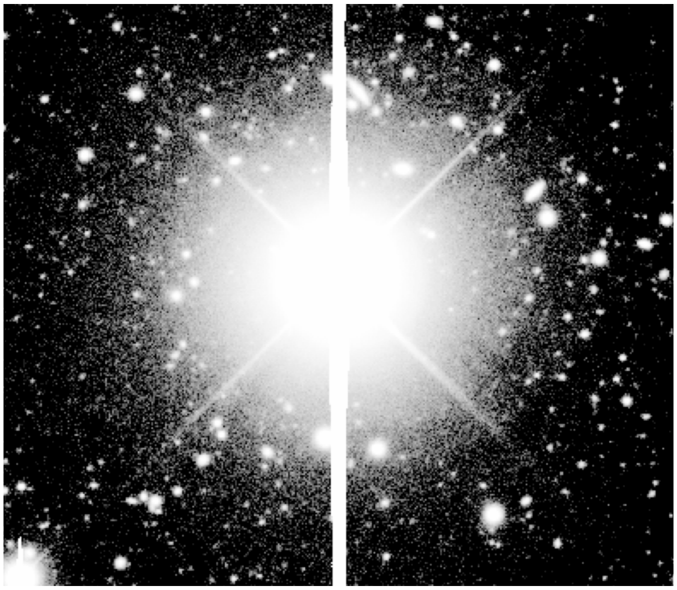

Fig. 2. Example of bleeding of a start on a CCD. Notice the large vertical line running down the image which has reached saturation [6].

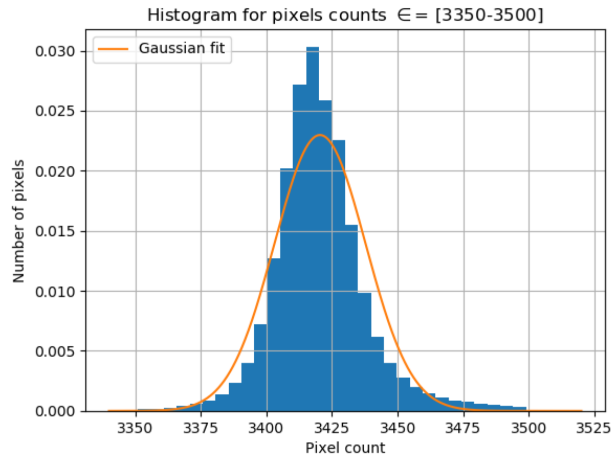

Fig. 3. Example of the distribution of pixels that made up the background noise. Notice the expected Gaussian distribution.

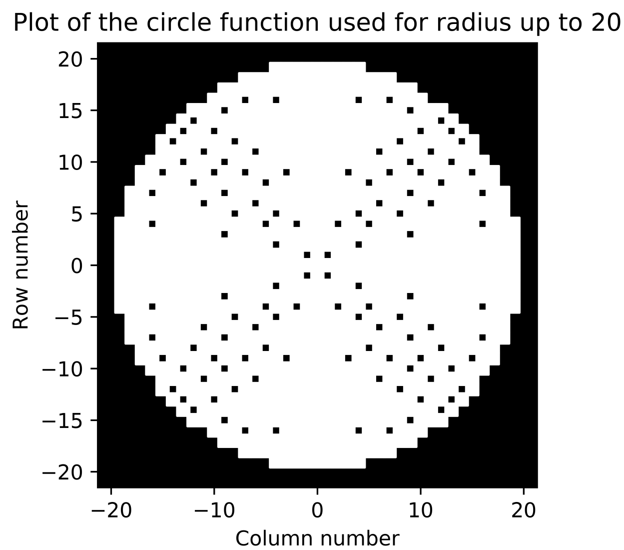

Fig. 4. Example of the circle function. It can be seen clearly in black the missing points that the circle function has missed.

III. Method

This image processing program consisted of function sorted through all the data points with brightness values above some local value and stored them in ascending order. This acted as the starting point for searching for galaxies, staring each time from the highest brightness pixel, and estimating each galaxy as circle. The radius was decided by choosing a cut off brightness value of pixels. This was repeated until all the pixels of interest had been dealt with, and a list of every galaxy with its brightness, pixels and position all recorded. The program was coded and tested in Python.

To start, it was first needed to determine which pixels from the plot were galaxy pixels, or pixels that were not noise. This was done by taking all data and splitting it in to 10 cells wide and 29 cells long. For each cell, a mean and standard deviation was taken, and from this anything pixel that had a brightness above two standard deviations of the mean was taken as a pixel that could be part of a galaxy. This was repeated for every cell, and allowed us to account for the variable noise in the image. A sample histogram of the noisy pixels is shown in Fig. 3.

These pixels were then sorted into descending order, in that idea that the centre of galaxies were likely to be the brightest. For this brightest pixel, an expanding circle is drawn around the centre, with the average brightness of each ring plotted as a function of radius. Once this average value dips a critical value, the radius of the galaxy is found, and the brightness of the galaxy can be found. The pixels within the galaxy can be removed from the list of galaxy pixels, and the next brightness point will follow the same procedure until there are no pixels left.

At first a ”snaking algorithm” was to be used which would spiral outwards from the centre pixel in order to find all the galaxy pixels, however this was found to be more computationally intensive, and required a more difficult condition of when to stop. A ”square function” was also considered very similar to a circle function but given the elliptical nature of galaxies it was decided a circle function is a better estimate. Of course the best approximation would be an elliptical function, however this requires two axis and also an angle to describe the orientation of the galaxy which is more computationally intensive and beyond the scope of this short program.

The circle function was based on the midpoint algorithm, and would produce approximations of circles on integer coordinates or cells [4]. However this was not without its faults, as shown in Fig. 4. A series of radii are drawn with the function but very clearly there are gaps where no points have been missed. This was dealt within the analysis, but also produced issue in terms of non-deleted pixels being left behind not picked up by the circle function. This was fixed by not allowing one pixel galaxies to be collected.

The final form of the data would be a list of every galaxy, its position, radius and brightness from which a magnitude count plot could be produced as discussed later in the results.

Before processing the image, the image was edited as shown in figure. This was because very clearly in the centre there is a start with bleeding, but clearly galaxies are what should be detected. A more sophisticated way to deal with this would be to class objects as either stars and galaxies and therefore decide whether an object should be counted or not, however given the size of the image the star and other bleeding was cut out manually. Another change that was made was the edges of the images are cut because there are regions of the image which appear to be manually blacked out and a very visible border that seems inconsistent with the rest of the image. Fig. 5 shows the change made to the image.

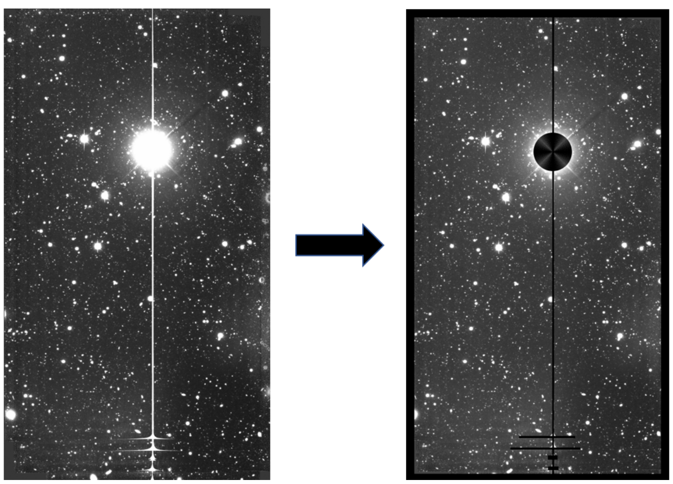

Fig. 5. The original image that was used and the image that was edited and analysed. The changes made were the blocking out of the big star, the bleeding from the stars and also the border around the edges [7].

IV. Results and Discussion

From the image, 1112 galaxies were found, and the gradient of the magnitude count plot yielded 0.42 ± 0.01. This graph is shown in Fig 6. with the line fit to the linear region shown in orange.

The fitted line is only plotted for the lower magnitudes of galaxy. This is because it was evaluated that the program performed best for brighter galaxies since less bright galaxies were harder to tell apart from the noise so it was decided that this region was better tot take data from.

The error was taken by considerations of the error in the magnitude. By taking counts with the magnitude plus the error and the magnitude minus the error, this variation was taken as the error in the counts. Using these errors for the weighted fit, an error in the gradient of the graph could be produced.

Comparing this gradient to the expected theoretical value from 3, and to the the paper by Yasuda which yielded a value of 5.3, the value of the experiment is small [78. There are a few possible reasons for this.

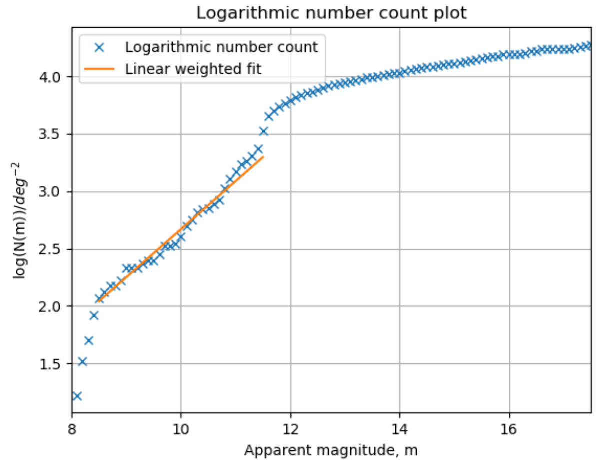

Fig. 6. Graph of the log count of galaxies against the apparent magnitude. In orange is the weighted fit with the points plotted in blue.

Firstly, as derived in the theory, 3 assumes a normal Euclidean volume, when in fact General Relativity applied to the whole universe suggests that this may not be the case, and volume may change in a greater rate that it would for a normal Euclidean volume. However from this small sample of data it is not possible to say which much confidence that this is responsible for the deviation seen.

Secondly, since the speed of light is finite, information from different distance correspond to difference times and therefore is a not a snapshot of the universe in the way that is treated. This would explain why the gradient is different, but not specially why the value is too small.

Thirdly, there is incompleteness in that some of the sources are too faint, and therefore the data set is complete. This is a limitation of the equipment but it can be said that there is less galaxies than expected as the distance increases.

Fourthly, it is possible that some stars were counted as galaxies. While some of the starts were removed, and could be identities by their bleeding and circular shapes, it cannot be ruled out the existence of stars in the data.

Lastly, it was estimated that teach galaxy was circular in shape but very clearly this is not true and therefore the brightness is greater than it should be. This is partially addressed in that fact that the background signal was accounted for, however whether a better way to deal with this would have been to detect the galaxy is elliptical forms.

It is worth discussing how the circular function performed. Fig. 7 shows how the average brightness of the pixels varied as a function of the radius. The results is what is expected with the count decreasing smoothly until some noise is reached. Fig. 8 shows how the cumulative brightness changes with pixel count. As expected it increasing smoothly until it saturates and some constant gradient is reached.

Each Galaxy was searched for a maximum radius of 20, and by inspecting the image by hand this seemed to be reasonable assumption since galaxies had a radius of 20 or less and the smaller the searched radius the quicker the program would run. However it was repeated for only a radius of 12 pixels and 1105 galaxies were found which is only 22 less. This is very small compared to the number of galaxies and justified our assumption that most galaxies had a radius less than 20 pixels.

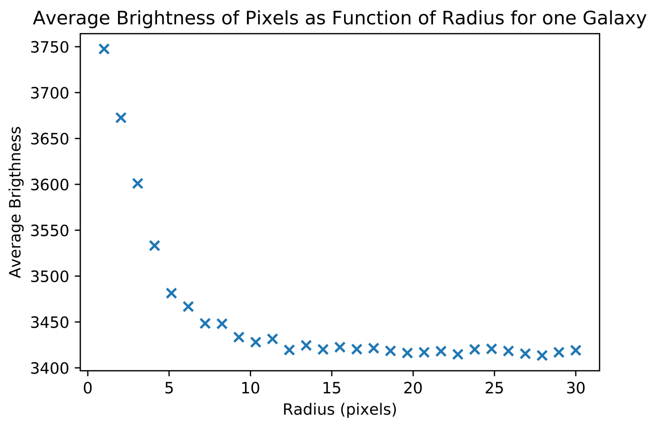

Fig. 7. Graph showing how the average brightness of the pixles varied with incrasing radius.

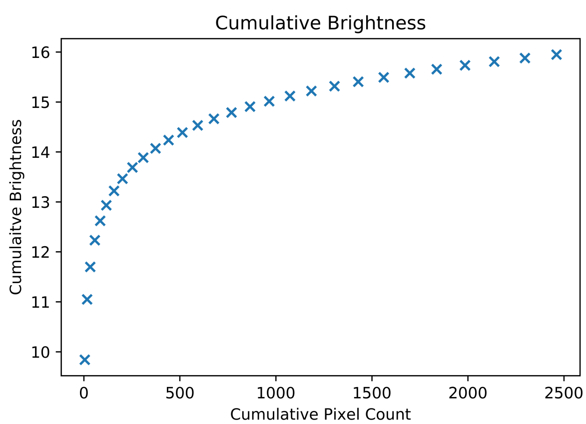

Fig. 8. Graph of the cumulative brightness of the pixels.

V. Conclusion

The aim of this experiment was to produce a program that could read an astronomical image and identity the galaxies, as well as information about its composition and brightness. From this, magnitude count plot could be produced and be compared to other standard experiments as well as the expected theoretical value.

This was done by searching the image for the brightest pixels above some critical value. The brightness pixel was taken as the centre of a galaxy, and it was approximated as a circle. The points of this galaxy were recorded, and each pixel was iterated through until all the pixels were collected. The program could then return a list of all the galaxies and their respective positions, compositions and brightness. The program found 1127 galaxies and produced a gradient of the magnitude count plot of $0.42/pm0.01$, compared to the expected 0.6.

This experiment was successful in that it was able to detect account for noise from the image locally, which meant it could more accurately distinguish noise from objects of interest. And because of this even though the galaxies were approximated as circles, any empty pixels included should have not have contributed anything since they were just noise.

Failures include the fact that any bleeding or stars had to be dealt with manually. This was okay for dealing with a single image, however anything more would required some kind of implementation of distinguishing between a star and a galaxy, and also the effect of bleeding. Since the program was written in Python, it runs slower than if it had been written in an lower level language, which again is a problem for scaling and not this particular piece of data. And as discussed a better approximation would be to take the galaxies as elipsses and not circles, which would requires significantly more work since two radii are required as well as an angle.

VI. References

1) Jones MH(H, editor., Lambourne RJ, editor., Serjeant S, editor., Cayless A, author. An introduction to galaxies and cosmology. Second ed. Cambridge: Cambridge University Press; 2015.
2) Dr. D.L. Clements, N. Skrzypek, Y Unruh. Astronomical Image Processing. A1nbsp; : Imperial College London; 2019.
3) Galactic onionnbsp; Available from: https://www.spacetelescope.org/images/potw1519a/.
4) Simple Circle Algorithmsnbsp; Available from: https://www.cs.helsinki.fi/group/goa/mallinnus/ympyrat/ymp1.html.
5) Longair MS, 1941-. Galaxy formation. 2nd ed. Berlin: Springer; 2008.
6) Walker G. Astronomical observations : an optical perspective. Cambridge: Cambridge University Press; 1987.
7) Images. Available from: http://www.spitzer.caltech.edu/images.
8) Naokoi Yasud. Galaxy Number Counts from the Sloan Digital Sky Survey Commissioning Data. 2001