UNIVERSITY of CALIFORNIA

SANTA CRUZ

**IDENTIFICATION OF
FAINT GALAXIES AT HIGH REDSHIFT**

A thesis submitted in partial satisfaction of the
requirements for the degree of

BACHELOR OF SCIENCE

in

ASTROPHYSICS

by

Beth Nordeen

8 June 2011

The thesis of Beth Nordeen is approved by:

_____   _____
Garth Illingworth   Adriane Steinacker
Advisor   Theses Coordinator

_____
David Belanger
Chair, Department of Physics

Abstract

Identification of

Faint Galaxies at High Redshift

by

Beth Nordeen

It is widely accepted that the universe began in a hot, dense state. As it cooled, ions condensed into neutral hydrogen atoms. After the universe was ~1 billion years old, most of this hydrogen had become reionized. The beginning of this reionization era is coincident with the formation of early galaxies. Could early galaxies have emitted the UV light required to reionize the universe? How do we *find* them? Here we present a novel method for detecting light from very faint (image noise level), early galaxies at a redshift range of 4 to 5 ($z \sim 4 - 5$) in deep images. Galaxies in this redshift regime are more accessible in current images, allowing us to develop this method for use with future images that probe earlier times. The new method uses an image we build called the *redshift factor* – an image that has a value (RSF) of ~1 at locations of galaxies at $z \sim 4 - 5$, and a value of ~0 at the locations of galaxies at $z < 4$. Ninety percent of galaxies we detected were identified as being at the correct z by the new method, and <0.1% of galaxies at $z \sim 4$ were missed by the new method. Surface densities of RSF-selected, high-z sources calculated by us were similar to surfaces densities from earlier work by another team, and were slightly more robust at brighter magnitudes. We conclude that the new method produces a robust sample of high-z galaxies. In the future, the redshift factor could be used to select individual pixels that have light from high redshift. This could provide a measurement of light from high-z galaxies that are not detected as individual sources due to their low flux levels, thus increasing the amount of information extractable from deep images. When this technique is eventually extended to probe $z \sim 6 - 20$ (reionization era), this summed light could be compared to the amount of light required to reionize the universe in order to determine whether early galaxies are candidates for producing this light.

Contents

List of Figures

List of Tables

To

my mother and father,

who gave me this opportunity to reinvent myself.

Acknowledgements

I would like to thank the people who cheerfully and patiently advised me on this project, which involved many layers of new knowledge and skill. They are: Rychard Bouwens, Valentino Gonzalez, Bradford Holden, Garth Illingworth, and Daniel Magee. I would like to extend a special recognition to Rychard Bouwens, who taught me the scientific principles behind this work. His guidance was invaluable. I would also like to recognize Bradford Holden, who tirelessly guided me in the eleventh hour to the end point of this thesis. Many thanks to Garth Illingworth, who worked hard to make sure I continued to have financial support throughout the duration of the project. I am indebted to Adriane Steinacker, who welcomed me and encouraged me as a returning student. I did not find such openness to, and support of, my desire to pursue an astrophysics degree at any other institutions I considered. I would like to thank Chris Myers at Cornell for his "Scipy array tip sheet", which condensed many pages of Scientific Python documentation into one handy webpage. Anthony Aquirre's General Relativity was the pinnacle of coursework I did at UCSC: an intensely spiritual experience! Piero Madau paved the way for that one by showing me the basics of my entire reason for showing up at UCSC in the first place: cosmology. All of these important people have allowed me to burst through the academic glass ceiling I was feeling before returning to school.

1 Introduction

Asking questions about the history of the universe is exploring our origins. Key cosmological observations have served as signposts along this journey to understanding. Figure 1.1 shows a visual history of these events, which are summarized below.
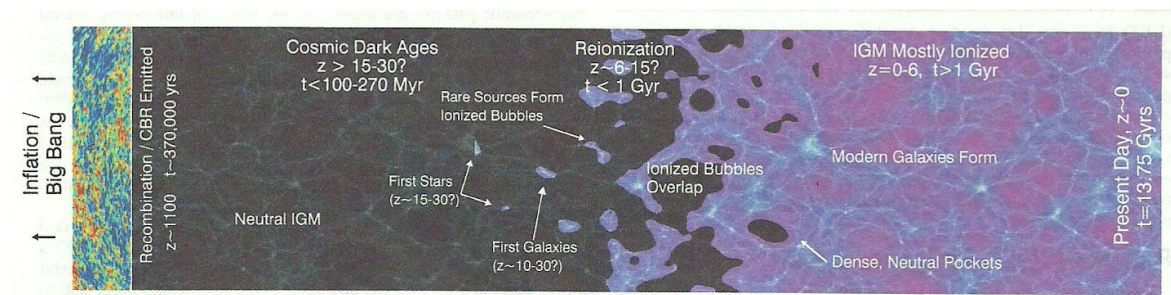
Figure 1.1: Cosmic history from the Big Bang until today, with an emphasis on the reionization era [Robertson, et al., 2010].

In 1965, evidence that the universe began in a hot, dense state was captured in the form of the cosmic microwave background (CMB) [Penzias and Wilson, 1965]. This blackbody radiation was released from the expanding infant plasma universe when it cooled down enough for matter, mostly non-ionized atomic hydrogen gas (HI gas), to condense. After condensation, light was free to make its way across space and time to meet varied fates. Many photons scattered, and a few have ended their travels at the detectors in modern telescopes. After this release of photons, the continually expanding universe entered a time period called the Dark Ages — stars had not yet formed.

Until 1965, it was assumed that, during the time of galaxy formation and at all times

thereafter, hydrogen existed as neutral HI gas. However, studies of spectra from the quasar 3C 9 revealed a compelling phenomenon: the hydrogen gas in the intergalactic medium was ionized at $z \sim 2$ [Gunn & Peterson, 1965]. Further research showed that the onset of this event was coincident with early galaxy and star formation after the Dark Ages ($z \sim 20$), and that by the time the universe reached an age of $\sim$1 billion years ($z \sim 6$), the intergalactic medium (IGM) was almost completely ionized [Barkana & Loeb, 2001]. What caused this reionization of the universe? To verify a connection between early galaxies and the reionization of the universe would be a great leap in understanding galaxy formation and its effect on the universe.

Ionization of HI gas requires photons with energies of 13.6 eV, corresponding to a wavelength of 912 Å. Hydrogen can also absorb light with a wavelength of 1216 Å, which facilitates a transition of its electron from the ground state to the first excited state without ionization occurring. Could stars in early galaxies have been a source of these UV photons? As galaxy density increased in the early universe, so did the density of photons produced by stars in the new galaxies, and many of these photons escaped into the universe. They were able to ionize the HI gas in the IGM, but were they plentiful enough at the time of reionization to accomplish this immense task? At present, the answer is still uncertain. To find out, astronomers must estimate how many of these photons were produced by early galaxies, and, of those, how many were able to escape and ionize the HI gas. An important preceding step is correctly identifying galaxies as existing at appropriate times and distances so they may be counted as photon contributors. This thesis describes a novel method for identifying faint, early galaxies in deep images when there is a low signal to noise ratio. To develop and calibrate this method for investigating earlier times, we test it on a redshift regime for which we currently have deep, faint data. Chapters 2 and 3 present a current technique for identifying distant, early galaxies in deep images and the results of its employment on test data. A detailed description of the new technique can be found in Chapter 4, and the outcome of its employment on test data is presented in Chapter 5. A discussion of the results is presented in Chapter 6, followed by conclusions in Chapter 7.

2 Lyman-Break Galaxy Search Technique

As light travels through the IGM from a distant galaxy toward our galaxy, its wavelength increases due to the expansion of the universe. This causes the light to appear redder than when it was emitted. The longer the light has been traveling, the redder it appears when detected. The factor associated with the stretching of the wavelength is called *redshift*, or "*z*". The more the wavelength has been stretched by the expanding universe, the higher the value of z. See Figure 2.1. If z is measured accurately, the age of the galaxy when the light was emitted can be estimated.
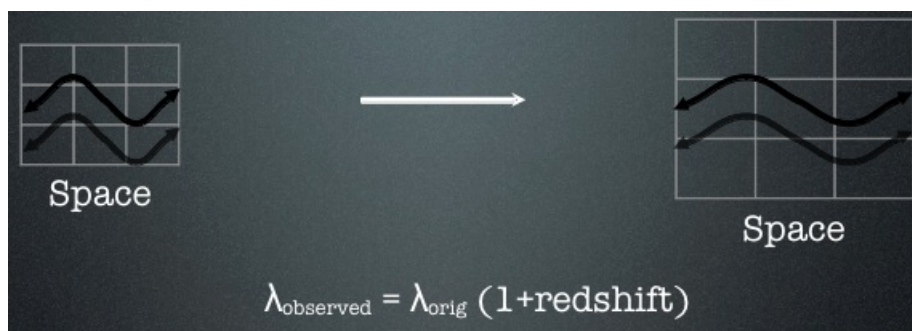
Figure 2.1: Redshift. As light travels through expanding space, its wavelength increases, causing it to appear redder. Image courtesy of Rychard Bouwens.

Finding the redshift of an early galaxy relies on the abundance of HI gas in regions that light from this galaxy may travel, because HI gas absorbs light at 912 Å or 1216 Å. This absorption causes the appearance of distinct features in spectra of these galaxies. There is HI gas in the atmospheres of the stars that make up distant galaxies, and in the interstellar space in younger galaxies. There are also great clouds of it in the intergalactic space, as well as a relatively thin distribution of it

filling the space between these clouds [Giavalisco, 2002] contributing to the IGM. Light coming from a star in a distant galaxy that is emitted at either 912 Å or 1216 Å can be absorbed by HI gas in the the stellar atmosphere before it even leaves the parent star. However, high energy light with wavelengths shorter than 912 Å or 1216 Å , along with other light that is not at these wavelengths, can escape. If it takes a path such that absorption by interstellar HI regions in the parent galaxy is avoided, this high energy light can begin to make its way toward our galaxy.

As it travels through expanding space, however, light that has wavelengths shorter than 912 Å or 1216 Å will be redshifted to these wavelengths and can then also be absorbed locally by incident clouds of HI gas. This can happen at any point on the light's path when the light is redshifted to these favored wavelengths and there is HI gas present. The result is that a large chunk of flux in wavelengths blueward of 912 Å (at the location of the emitting galaxy) will be missing from the putatively unattenuated rest-frame spectrum of the distant galaxy when finally measured. This discontinuity is called the *Lyman break*. There is also noticeable attenuation of the continuum blueward of the 1216 Å line, giving the spectrum a "staircase" appearance [Madau, 1995]. Galaxies whose spectra have this feature are called *Lyman break galaxies* (LBG's). Figure 2.2 shows an example spectrum of a LBG at $z \sim 4.5$.
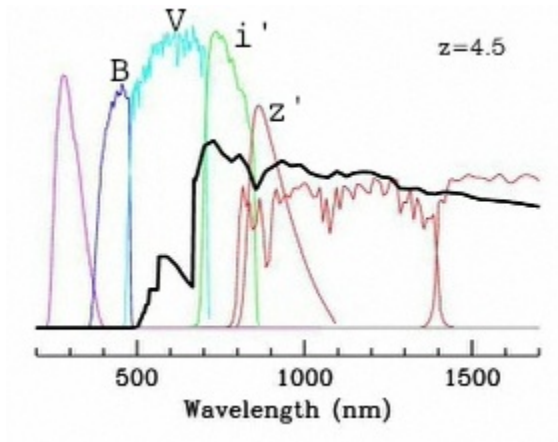
Figure 2.2: An example spectrum (black line) of a Lyman break galaxy at $z \sim 4.5$ shown over the transmission bands of each filter used. Note the abrupt drop of flux in the spectrum blueward of 500 nm, and the "staircase" feature at 700 nm. Figure courtesy of Rychard Bouwens.

The key is that the position of this break in the spectrum will be redshifted away from 912 Å according to the redshift of the parent galaxy. The attenuated spectrum (black line) in Figure 2.2 clearly shows the Lyman break at wavelength at 500 nm (5000 Å), and the flat continuum at bluer wavelengths. An unattenuated spectrum would be observed if there were no absorption by HI gas. The observed location of the Lyman break at 500 nm is related to the redshift of the galaxy in the following way:

$$redshift = \frac{\lambda_{absorbed} - \lambda_{emitted}}{\lambda_{emitted}} = \frac{5000 \text{ Å} - 912 \text{ Å}}{912 \text{ Å}} \approx 4.5 \tag{2.1}$$

Also in Figure 2.2, the colored curves represent transmission bands of each filter used. Notice that 500 nm (5000 Å) coincides with the cutoffs between two of the filters: the B- and V-bands. Galaxies at $z \sim 4.5$ are very dim in the B-band and bright in the V-band, and therefore "drop out" of images in the B-band. Such galaxies are referred to as *B-band dropouts*.

In order to quantify the severity of the Lyman Break, we use ratios of source flux in the different passbands, or *colors*. Colors are expressed as differences in magnitudes, and the magnitude of a source is related to the logarithm of the flux. Magnitudes used in our calculations are in the AB magnitude system. AB magnitude is defined by

$$AB = -2.5 \log (f_\nu) - 48.60 \tag{2.2}$$

where f_ν is the flux in ergs cm^{-2} s^{-1} Hz^{-1} [Oke & Gunn, 1983]. The color B-V, for example, would then be the difference between the magnitude of the source in the B-band (m_B) and its magnitude in the V-band (m_V). This is the ratio of the flux in the two bands, and thus would be calculated

$$m_B - m_V = -2.5 \log \frac{f_B}{f_V} = B - V \tag{2.3}$$

If the flux in the V-band is larger than the flux in the B-band, then the color B-V would have a value greater than zero. The larger this difference in flux, the higher the value of B-V. Colors having a value greater than zero are considered "red", and colors having a value less than zero are considered "blue". Sources that are B-band dropouts are red in B-V and blue in V-i. Plotting sources in

colorspace can help identify those sources with particular color characteristics. An example of this can be found in Chapter 3.

At $z \sim 5$ and greater, absorption due to electron transitions in HI gas requiring photons with a wavelength of 1216 Å become more important. In the rest frame of the observed galaxy, there can exist light redder than 912 Å but bluer than 1216 Å that, if it escapes the galaxy, can be redshifted to 1216 Å as it travels. This light may also be absorbed by HI gas in the IGM. Consider the attenuation of the spectrum of a galaxy at $z \sim 7$ shown in Figure 2.3. The Lyman break at 1000 Å is associated with the rest frame 1216 Å line in the spectrum, not the 912 Å line. The tiny lump in the continuum at ~ 800 Å in Figure 2.3 is what remains of light with rest frame wavelengths between 912 Å and 1216 Å. Compare with the larger "hump" between 500 Å and 700 Å in Figure 2.2. As it becomes redshifted to 1216 Å, the light is absorbed by HI gas in the IGM. At these higher redshifts, the dropout effect in filtered images must be associated with the 1216 Å line in calculations of redshift instead of the 912 Å line:

$$redshift = \frac{\lambda_{absorbed} - \lambda_{emitted}}{\lambda_{emitted}} = \frac{10000 \text{ Å} - 1216 \text{ Å}}{1216 \text{ Å}} \approx 7 \qquad (2.4)$$
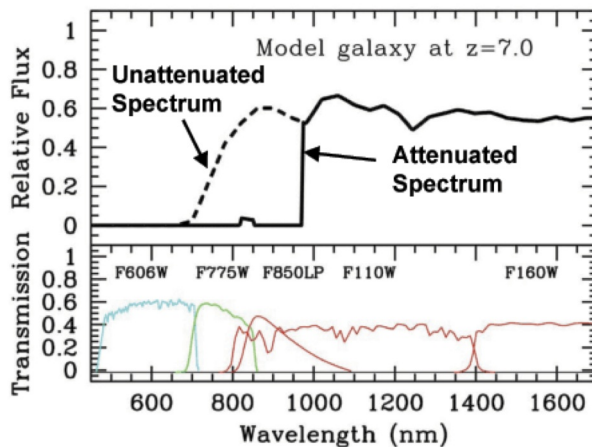
Figure 2.3: Example of a galaxy at $z \sim 7$ showing attenuated and unattenuated spectra. This galaxy would be a z′-band dropout. Image courtesy of Rychard Bouwens.

The technique of finding the redshift of galaxies this way, first used in 1990 [Guhathakurta, 1990],

is known as the *dropout method*. Figure 2.4 shows a correlation between redshift and the filter where the galaxy drops out of images: the redshift is plotted on the horizontal axis, and the throughput functions of each filter are plotted to show that objects at the corresponding redshift will drop out images taken through that filter. For example, an object at $z \sim 8$ is visible in images taken through the J filter, but very dim or not visible in images through the z filter or those to the left of the z filter (U, B, V, and i in Figure 2.4).
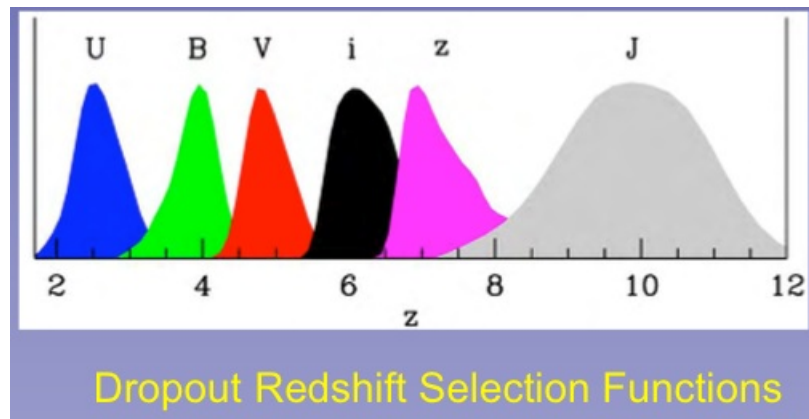
Figure 2.4: This figure shows the correlation between redshift and drop out filter. Figure courtesy of Rychard Bouwens.

3 Identifying Bright High-Redshift Galaxies in the Hubble Ultra Deep Field

3.1   Data Set

In order to estimate the amount of light coming from the first galaxies, the galaxies must first be detected. In 2004, the Hubble Space Telescope (HST) aimed at a $202'' \times 202''$ area in the constellation of Fornax for 400 orbits. This is the Hubble Ultra Deep Field (HUDF) [Beckwith, 2006]. Its location in the night sky is shown in Figure 3.1.

Eight hundred 20-minute exposures taken with the Advanced Camera for Surveys Wide Field Camera (ACS/WFC) through four different filters were combined to make this stunning view of distant galaxies seen in Figure 3.2. The long exposure times allowed extremely faint galaxies to be seen for the first time. The observations were made over a four month period, from September 2003 to January 2004.

Each of the four filters used to make the HUDF is transparent to light in a specific range of wavelengths, or, *passband*. Table 3.1 presents details on the filters, and Figure 3.3 shows the throughput function[1] for each of the filters. The throughput is a measure of sensitivity of the ACS/WFC. The instrument has 45% throughput at 700 nanometers (including the HST optical telescope assembly (OTA)) and is optimized for surveys in the optical and near-infrared to search

---

[1]http://acs.pha.jhu.edu/instrument/photometry/
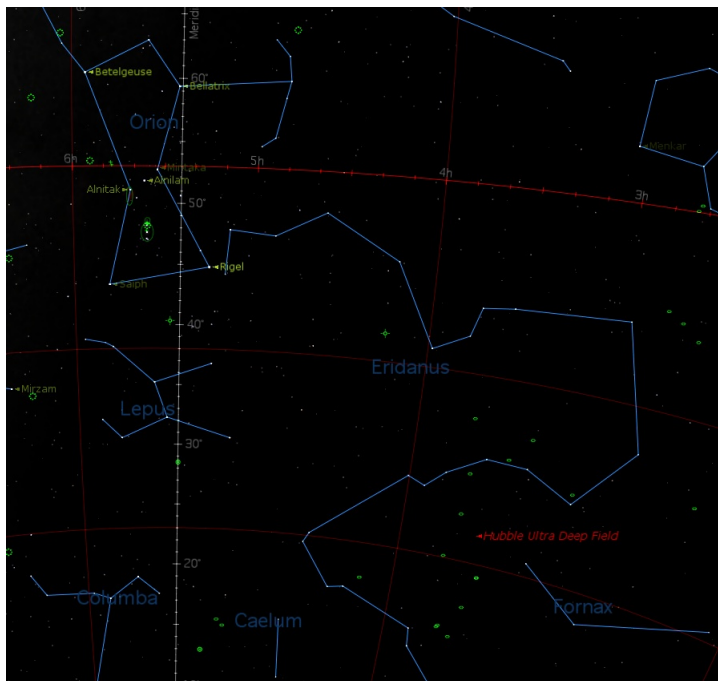
Figure 3.1: The location of the Hubble Ultra Deep Field in the constellation Fornax, indicated by the red label in the lower right corner of this 55-degree-wide image. The 202-arcsec by 202-arcsec field was chosen for a 2004 survey to sample galaxies in the early universe.

for galaxies and galaxy clusters in the early universe[2].

Table 3.1: ACS filter specifications. Data from STScI ACS Handbook, Cycle 18.

Filter Name	Central Wavelength (Å)	Width (Å)	Description
F435W	4297	1038	Johnson B
F606W	5907	2342	Broad V
F775	7764	1528	SDSS i$'$
F850LP	9445	1229	SDSS z$'$

The four passband images used to detect galaxies using the dropout method (described in Chapter 2) were obtained from the Space Telescope Science Institute (STScI) website's Data Archives[3]. They are the result of dithered images that were combined using the Drizzle algorithm[4]. Each image is 10500 x 10500 pixels (420 Mb each), with a pixel scale of 0.03"/pixel, and oriented with North toward the top.They are in the Flexible Image Transport System (FITS) format.

---

[2]http://acs.pha.jhu.edu/instrument/overview/
[3]http://archive.stsci.edu/pub/hlsp/udf/acs-wfc/
[4]http://www.stsci.edu/hst/HST_overview/documents/multidrizzle/multidrizzleTOC.html/

Figure 3.2: The Hubble Ultra Deep Field. Credit: NASA, ESA, S. Beckwith (STScI) and the HUDF Team.

3.2 Procedures

To search for high redshift galaxies in these images, we first made a detection image. The detection image is a combination of images in passbands where $z \sim 4.5$ galaxies are visible: the V-band and the i′-band. Source Extractor (SExtractor) [Bertin & Arnouts, 1996] software was used to detect sources in the detection image. The locations of sources could then be mapped onto the B-band image, where the sources were very faint or absent. The criterium for significant detection in the detection image was that the ratio of a source's flux in the V-band to the error in that flux was greater than five (FLUX/FLUX_ERR > 5). SExtractor assigns a value of 99.0 to a source's magnitude if the detection is below the one-sigma limit, so we reassign a V-band magnitude of the one-sigma limit to these sources. We then use pixel weight maps to build a boolean mask selecting pixels with weights greater than half the maximum value of the weight map for each of the three passbands. The weight images are produced by Drizzle, and are the total exposure time
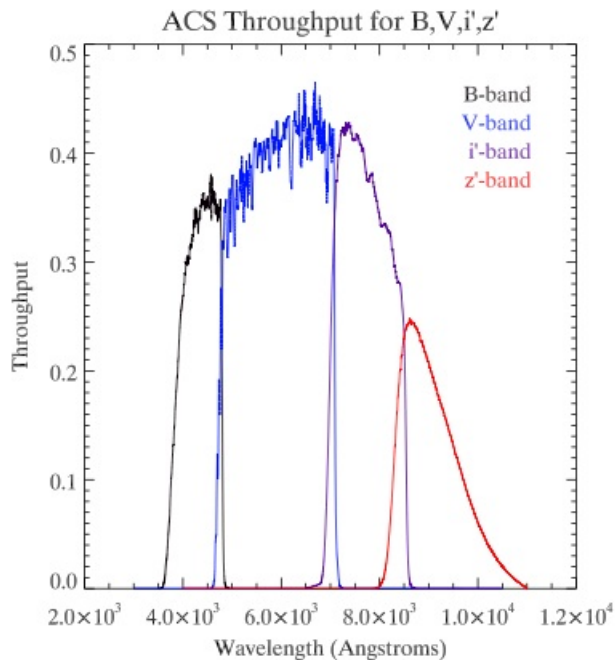
Figure 3.3: HST ACS/WFC throughput as a function of wavelength, on a scale of 0 to 1, where 1 corresponds to 100% transmission of light to the detector. The instrument has 45% throughput at 700 nanometers (including the HST optical telescope assembly (OTA)). Figure courtesy of Sara Ogaz.

for each pixel. We apply the boolean mask to segmentation map (an image produced by SExtractor indicating regions of pixels associated with sources in the HUDF) and select source regions that contain 80% "good" pixels. We called these "good" sources. We then compute the colors B-V and V-i$'$ and analyze the colors for each good source to determine if it is at $z \sim 4.5$. The details of calculating color are described in Chapter 2.

Figure 3.4 shows a flowchart outlining the image processing Python scripts, or *pipeline*, used to find $z \sim 4.5$ galaxies in the HUDF. Python programming language version 2.6.1 was used on a Mac Mini running Mac OS X Version 10.6.6. The Mini has a 2.66-GHz Intel Core 2 Duo processor and 2 GB of RAM.
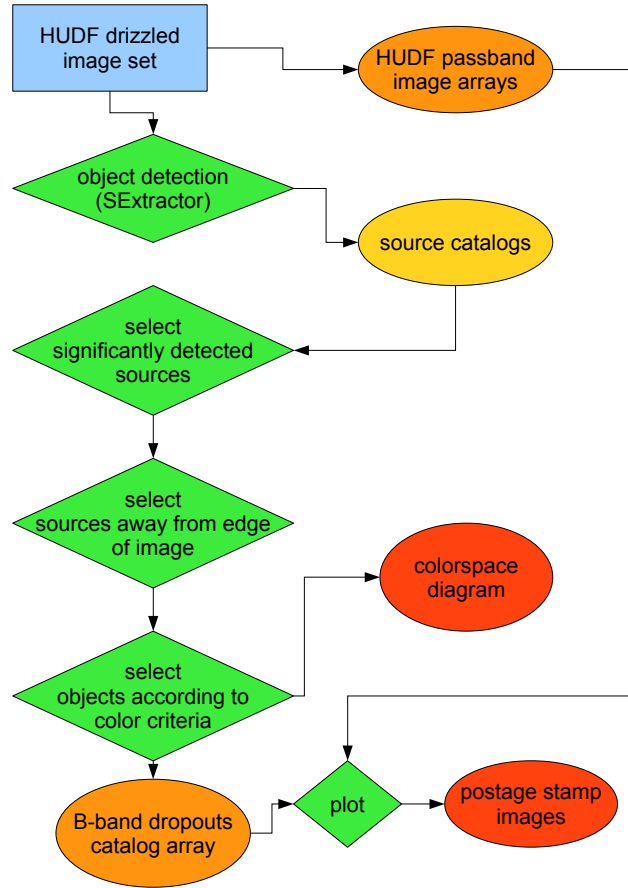
Figure 3.4: Dropout method image processing pipeline flowchart. This Python script selects $z \sim 4.5$ galaxies from the HUDF. It plots B-band dropouts in colorspace along with other objects significantly detected by SExtractor that are not near the edges of the passband images. It also produces postage stamp images of B-band dropouts in each of the four passbands.

3.3   Results

Figure 3.5 presents the 5768 sources significantly detected by SExtractor, in colorspace. Color criteria for a source to be considered a B-band dropout were B-V > 2.5 and V-i < 0.5. Of the 5768 sources, 41 had these criteria are represented by the red dots. Sources not fitting these criteria are represented by green dots. Motivation for these selection criteria is addressed in Section 3.4.

Figure 3.6 shows four zoomed-in passband images for five pseudorandomly chosen B-band dropouts from the set of 41. The postage stamp images clearly show the weak to nonexistent
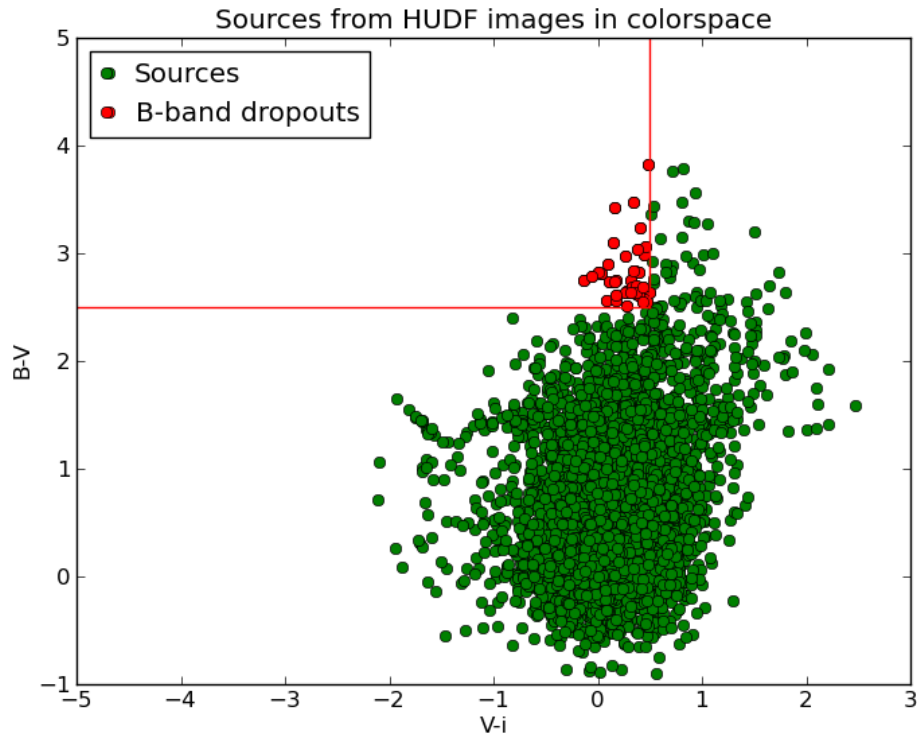
Figure 3.5: Significantly detected sources from HUDF in colorspace. The author's 41 B-band dropouts are indicated in red, and the red lines show the criteria for color selection. Sources not falling within the color criteria are represented by green dots.

contribution of these five in the B-band when compared with their appearance in the V-, i'-, and z-bands.

Figure 3.7 is a black and white image of source locations in the HUDF produced by SExtractor, called a segmentation map. White regions indicate pixels associated with significantly detected sources, and black regions are between sources. Highlighted on this map are B-band dropouts found by Rychard Bouwens (RB, red X's) [Bouwens, et al., 2007], and by the author (BN, green circles), independently. The sources in BN's set are the same 41 shown in red in Figure 3.5. Criteria for RB's selection of B-band dropouts were B-V > 1.1 and V-z' < 1.6 (V-z' is proportional to V-i'). These criteria are less conservative than those of BN, so, ideally, BN's set should be included in RB's set. Of the 41 detected by BN, 36 were also detected by RB, yielding an 88% overlap.
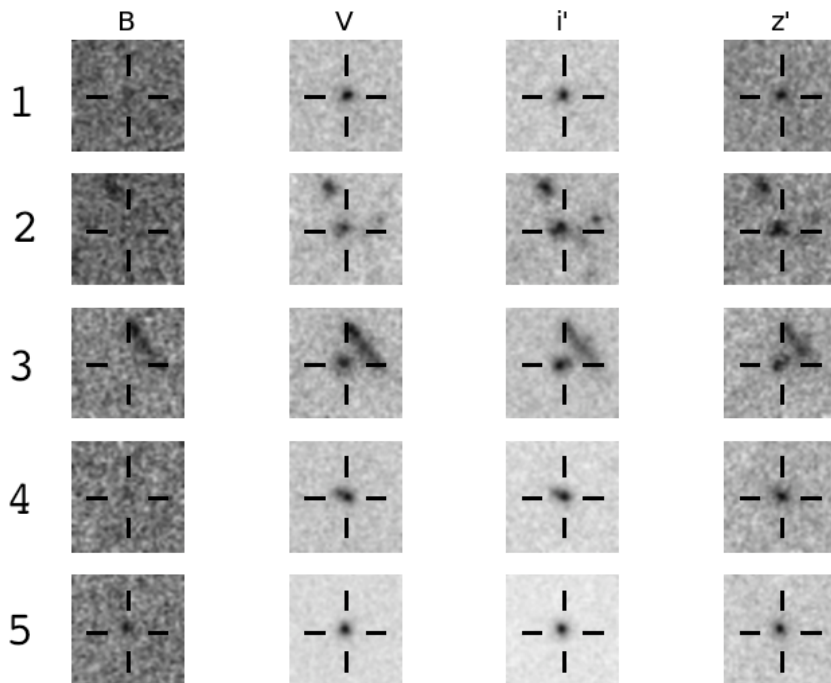
Figure 3.6: Pseudorandom sample of B-band dropouts from the HUDF. Each row represents one object and contains four images taken through each of the passbands labeling the columns. Notice the object is present in the V-, i′-, and z′-band, but very dim or absent in the B-band. This is a feature of objects at $z \sim 4.5$. In the fields of view of objects 2 and 3, there are other objects captured that are not at $z \sim 4.5$. They are therefore present in all four images.

3.4   Discussion

This traditional dropout method works well for bright sources with a high signal to noise ratio, but what about very faint sources? Are we missing them because their signal in the image is on the order of the noise level for the image? In the dropout method, we select objects based on their colors. Since the signal to noise ratio is 5 (a 20% error) in the detection image, and we are using combinations of images (colors, see Chapter 2) to analyze objects, the errors in the colors are even larger. Also, a single bright B-band pixel in a source region could cause that source to appear too blue to be considered a dropout. Even more fundamentally, we are dealing with objects that are, by definition, faint or absent in the B-band. This non-detection greatly increases uncertainty.

These issues could cause objects to be misidentified in colorspace. Is there a way to, in effect, boost the signal to noise ratio in hopes of detecting these important sources? Chapter 4 describes a new method of detection of faint sources at $z \sim 4 - 5$ that takes this into account. The set of 41 conservatively selected sources will be used to calibrate the new detection method. The method will then be tested on a set of $z \sim 4 - 5$ sources that have been selected using the traditional dropout method with the hope that these and more will be identified.
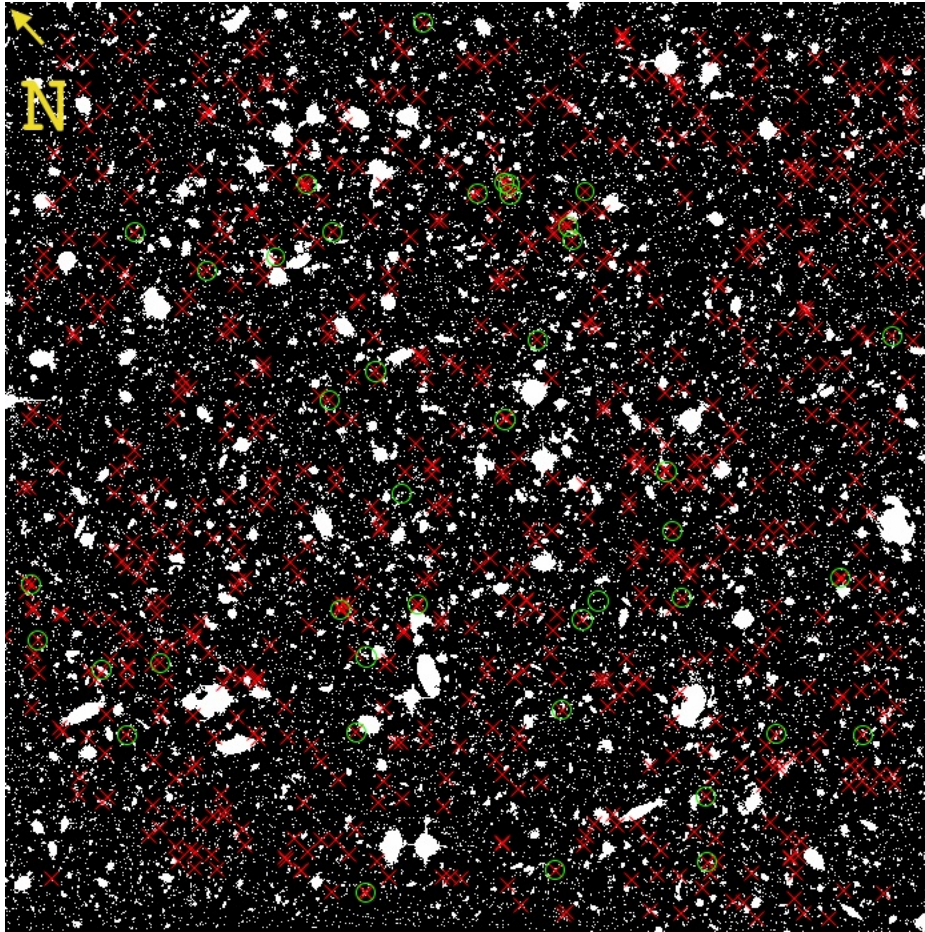
Figure 3.7: A comparison of B-band dropouts detected in the Hubble Ultra Deep Field by Rychard Bouwens (RB, red x's) and by the author (BN, green circles) independently. Criteria for RB selection: B-V > 1.2, V-z < 1.6. Criteria for BN selection: B-V > 2.5, V-i < 0.5. BN criteria are more conservative than RB criteria, so the set of 41 B-band dropouts selected by BN should ideally be included in the set of 730 detected by RB. Of BN's 41, 36 were also detected by RB: an 88% overlap. The 41 will be used to calibrate the new selection method described in Chapter 4. The black and white image is a map of the locations of objects detected by Source Extractor software.

4 New Procedure for Measuring the Light from High Redshift Galaxies

4.1   Theory

Because the traditional dropout method only works well for sources with a high signal-to-noise ratio, i.e., those detectable by SExtractor software, we want to develop a technique for identifying high-z galaxies that we can extend to very low signal-to-noise ratio. This is vital for counting the number of faint galaxies that contribute light at high z. The method we use to select galaxies at a particular redshift z relies on the idea behind the dropout method described in Chapter 2: galaxies at certain redshifts "drop out" of images taken through certain filters. We first create a detection image by summing the flux in passbands where the objects are visible. We then include flux through the passband where objects drop out and use the following equation to select galaxies at the desired redshift and greater:

$$\frac{(flux\ in\ detection\ image) - 2 \cdot (flux\ through\ dropout\ filter)}{(flux\ in\ detection\ image) + 2 \cdot (flux\ through\ dropout\ filter)} = redshift\ factor \qquad (4.1)$$

Formula 4.1 creates an image that has a value close to one at locations of galaxies at the desired redshift and greater, and has a value near zero at locations of galaxies at lower redshifts. This map is called the *redshift factor*. When this factor is used with a detection image, it selects galaxies at a particular redshift and greater.

For example, to select galaxies at $z \sim 4 - 5$ in the HUDF using this method, we make a detection image by adding the flux through the V and i′ bands. We then include flux through the B-band and use Formula 4.1:

$$\frac{(V + i') - 2 \cdot B}{(V + i') + 2 \cdot B} \tag{4.2}$$

The resulting redshift factor is an image that is $\sim$1 at locations of B-band dropouts (B $\sim$ 0), and is $\sim$0 where there is flux in the B-band. The image would also go to one at locations of objects at higher redshifts. V-band dropouts, for example, would be counted because the flux in the B-band and the V-band would be zero, but there would be flux through the i′ filter. The equation would still go to one. Figure 4.1 shows the theoretical value (RSF) of the redshift factor as a function of redshift for this example.
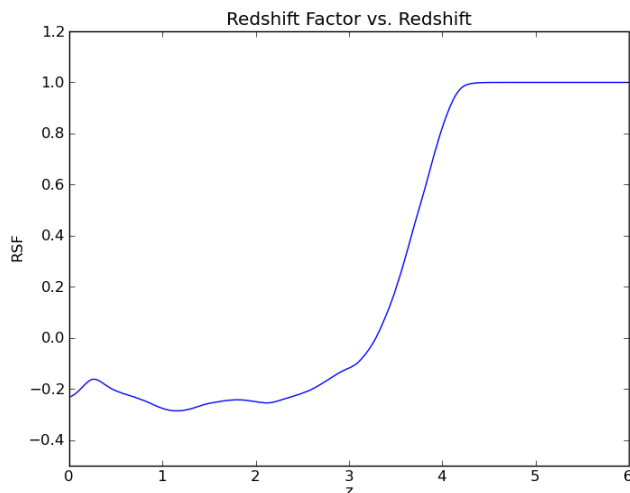
Figure 4.1: Theoretical value (RSF) of the redshift factor tailored for selecting $z \sim 4 - 5$ objects. This RSF function is built by using the flux of sources at $0 < z < 12$ in B, V, and i′ filters in Formula 4.2. The RSF hovers around -0.2 for $z < 4$ objects, and jumps to 1 for $z \sim 4 - 5$ objects.

This RSF function is built by using the flux of sources at $0 < z < 5.5$ in B, V, and i′ filters in Formula 4.2. The RSF hovers around -0.2 for $z < 4$ objects, and jumps to 1 for $z \sim 4 - 5$ objects. Data for galaxies at $z > 5.5$ were not included in the building of this plot. When we apply a map

of faint sources to the redshift factor image, we can calculate a RSF for each source and therefore identify which sources are at $z \sim 4 - 5$.

4.2   Procedure

To test this method on the HUDF, we first detected good sources in the the B-, V-, and i′-bands using the same method described in Chapter 3. We then created Python scripts to accomplish the following:

- make a detection image by gaussian-smoothing the V- and i′-band images, then adding them together

- create a redshift factor image by putting gaussian-smoothed B-, V-, and i′-band images into Formula 4.2

- create a RSF catalog of good sources and their median RSF values using segmentation map of the detection image

- plot the sources in a RSF histogram, colorspace diagram, and RSF vs. redshift scatter plot

Figure 4.2 shows a flowchart outlining the new image processing Python script, or *pipeline*, used to find galaxies at $z \sim 4 - 5$ in the HUDF.

We can then choose a RSF threshold based on where the conservatively selected calibration set of dropouts fall in the histogram. The threshold allows us to create a catalog of high-z galaxies from the catalog of good sources generated by the pipeline.
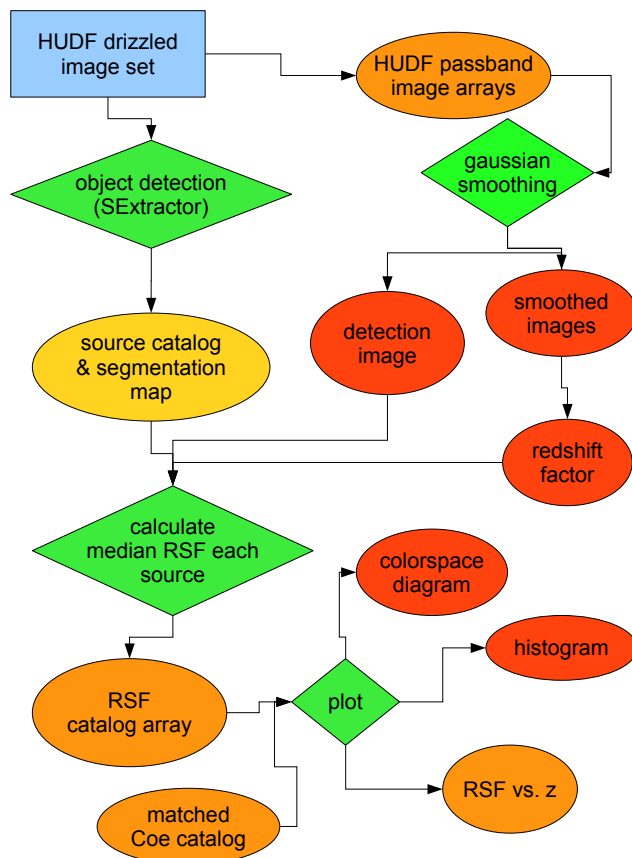
Figure 4.2: RSF method image processing pipeline flowchart. This Python script selects $z \sim 4-5$ galaxies from the HUDF. It plots objects in colorspace and uses bimodal color-coding to distinguish objects with a RSF≥ 0.7 from those with a RSF< 0.7 .

5 Results

Figure 5.1 is the histogram showing the distribution of 5370 of the 5768 sources in the RSF catalog among redshift factor values. These 5370 sources were chosen because they had V-band AB magnitudes dimmer than the minimum of the B-band dropouts, giving us a sample that is more uniform in characteristics other than the RSF. The red bins contain the 41 dropouts sources ("drops") discussed in Section 3.3 that were very conservatively selected by the traditional dropout method described in Chapter 2. The remaining 5329 "non-drops" are in the green bins. The drops have a mean RSF of 0.9 $\pm$0.1, and the non-ndrops have a mean RSF of 0.4 $\pm$0.2. We chose a RSF threshold of 0.7 to select galaxies at $z \sim 4 - 5$ because it included conservatively selected sources with a generous additional amount. We found that, including the drops, 749 sources of the 5370 had a RSF $\geq$ 0.7.

Figure 5.2 presents the 5370 galaxies in colorspace. The dots representing the galaxies are color-coded by their RSF. Red dots are drops which have RSF $\geq$ 0.7, and green dots are non-drops with RSF $<$ 0.7. We are now representing the 749 non-drops having RSF $\geq$ 0.7 with blue dots. Since we now know the colors of these sources, we can see how their colors compare with colors of sources at $z \sim 4 - 5.5$ selected by the traditional method with the following selection criteria: sources at $z \geq 4$ have B-V $>$ 1.1; sources at $z \sim 4$ have V-z$'$ $<$ 1.6 (V-z$'$ is proportional to V-i$'$), sources at $z \sim 5$ have V-i$'$ $>$ 1.2, and sources at $z \sim 6$ have V-i$'$ $>$ 2.8 [Bouwens, et al., 2007]. In Figure 5.2, sources with RSF $\geq$ 0.7 had a mean B-V value of 1.8 $\pm$0.5. Mean V-i$'$ was 0.4 $\pm$0.6, indicating a large spread in this color for these RSF-selected sources. There were 52 sources with V-i$'$ $>$ 1.2 and

zero sources with V-i$'$ > 2.8. These values are consistent with detections of sources at $z \sim 4 - 5$.

As a check, we matched the catalog of good sources to Daniel Coe's catalog [Coe et al., 2006], which contains independently determined photometric redshift values (photo-z's). Coe used BPZ software [Benitez, 2000] to estimate the z's of galaxies, which applies Bayesian probability to the photo-z estimation. Photo-z's are estimated redshifts based the colors (Chapter 2) of galaxies. Bayesian probability theory is centered not on frequencies as indicators of likelihood, but on how much is known about hypothesis. Bayesian probabilities are therefore dependent on other information. Benitez showed that using this theory gave more reliable results of photo-z estimation.

We successfully matched 4811 of our 5370 sources to sources in the Coe catalog. Figure 5.3 shows the median RSF factor for each source plotted against the photometric redshift determined by Coe (black dots). Plotted on top in blue is the theoretical curve from Figure 4.1. The shape of the observed curve agrees with the theoretical curve, although the RSF's at lower z tend to be greater than those in the theoretical curve. False selections (sources having a RSF ≥ 0.7 but are at $z < 4$) were found to be $\sim$10% of the sample. Possible missed selections (sources having a RSF < 0.7 but are at $z \geq 4$) were found to be $< 0.1\%$ of the sample.

Figure 5.4 shows surface density estimates as a function of i$'$-band magnitude: the two results shown are from the traditional dropout method (black dots) and the new method (red dots). The two sets of results are offset from each other by 0.1 magnitude for clarity. Errors are estimated by assuming that the error is the square root of the number counts in a bin (Poisson statistics). For small sample sizes, such as our low densities at the extremes of our i$'$-band magnitude range, this is no longer an accurate approximation and errors will be overestimated.
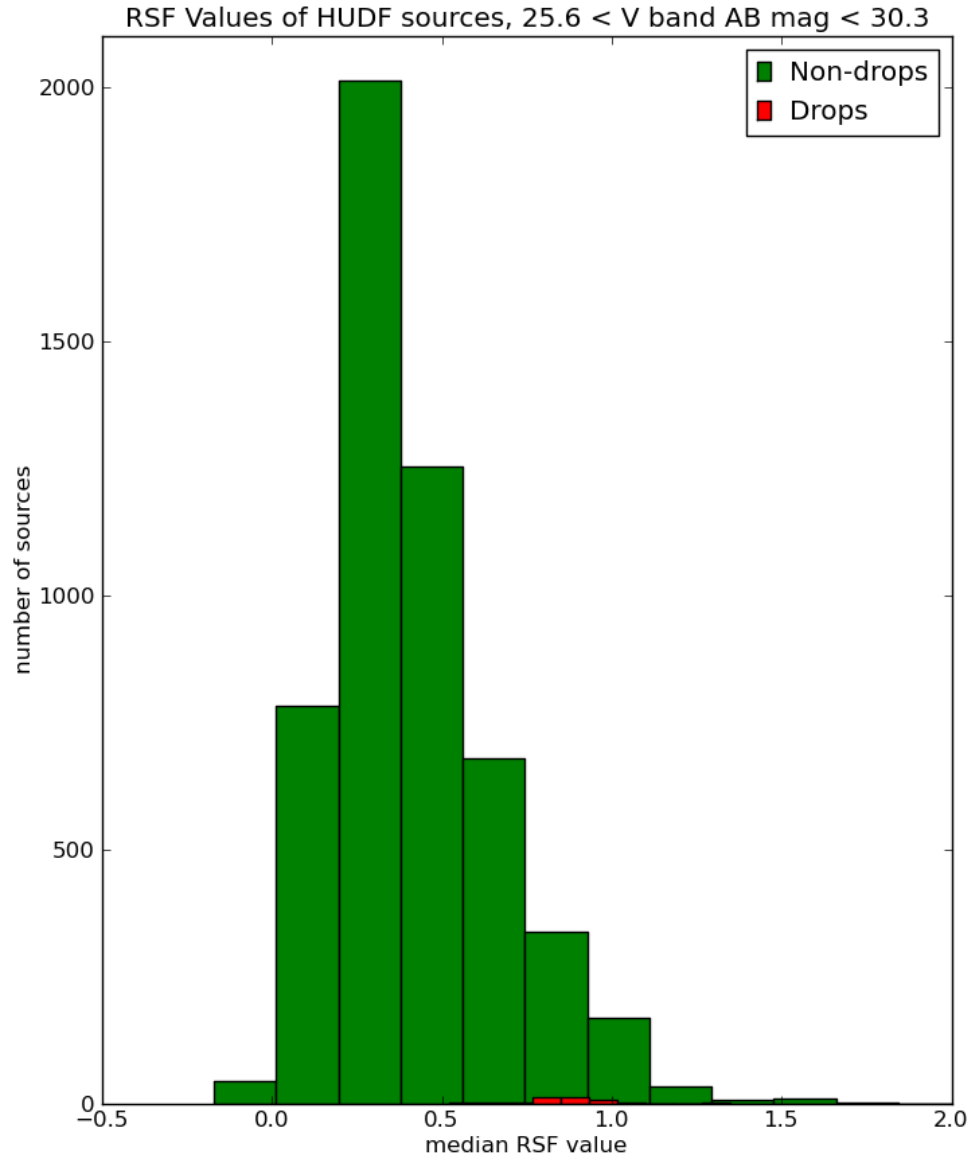
Figure 5.1: Histogram showing the number of sources assigned a particular RSF. The green bins hold galaxies that were not traditionally selected B-band dropouts (non-drops). The red bins hold traditionally selected B-band dropouts (drops). The drops have RSFs greater than 0.7. The non-drops in this figure with RSF > 0.7 could be galaxies at $z \sim 4-5$ that were missed by the traditional selection method.
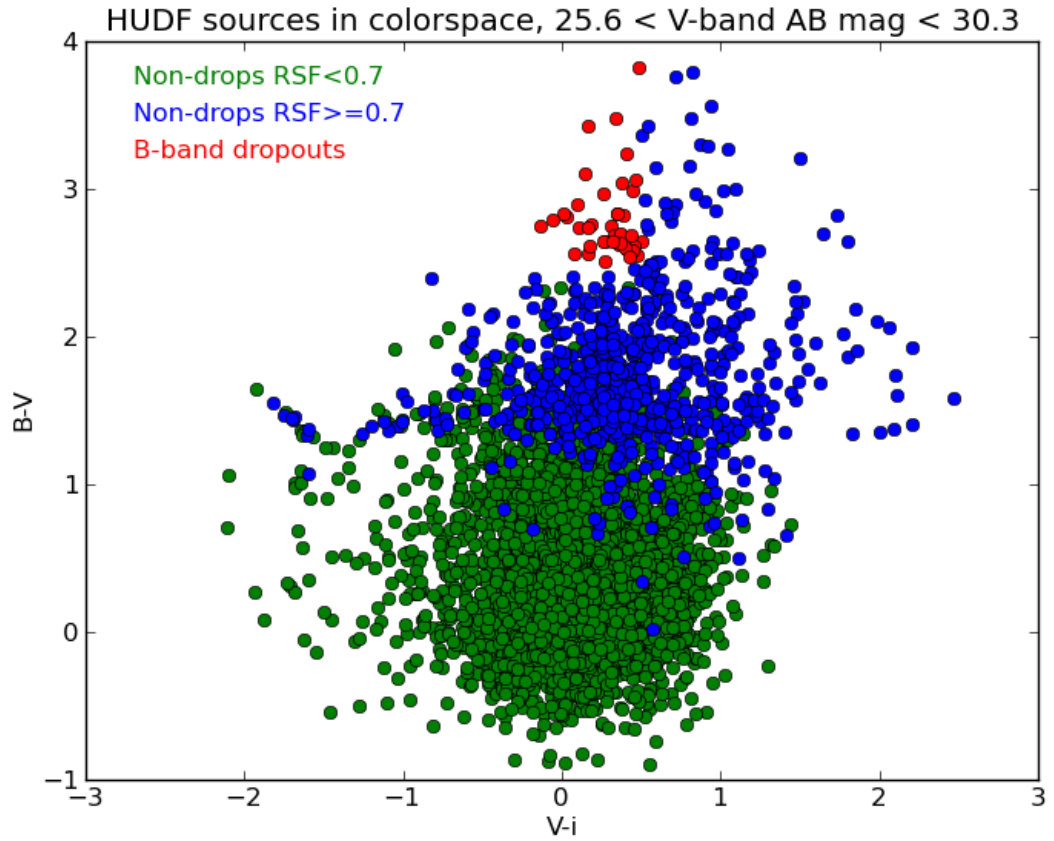
Figure 5.2: Significantly detected sources from HUDF in colorspace. Traditionally selected B-band dropouts are indicated in red, and non-dropouts with a RSF < 0.7 in green. Highlighted in blue are non-dropouts with a RSF ≥ 0.7. The blue non-drops could be galaxies at $z \sim 4 - 5$ that were selected by the new method.
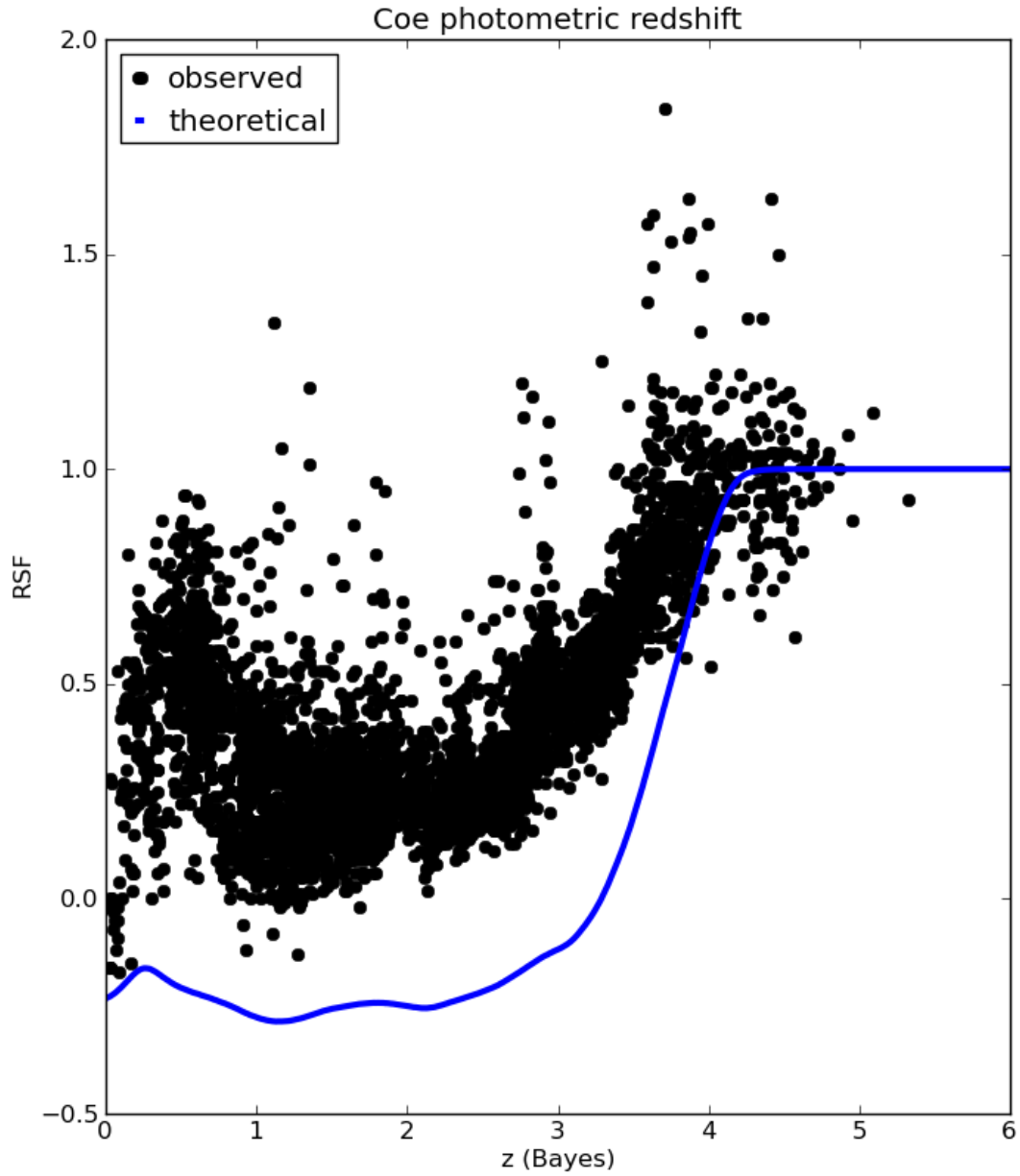
Figure 5.3: Photometric redshifts [Coe et al., 2006] of sources. Sources in our RSF catalog from the new method were matched with sources in a catalog having photometric redshifts. Matched sources are represented by black dots. The blue curve plotted over the matched sources is the theoretical redshift factor for each redshift from Figure 5.3. There is agreement between the two in shape, but the observed values are higher than the theoretical values.
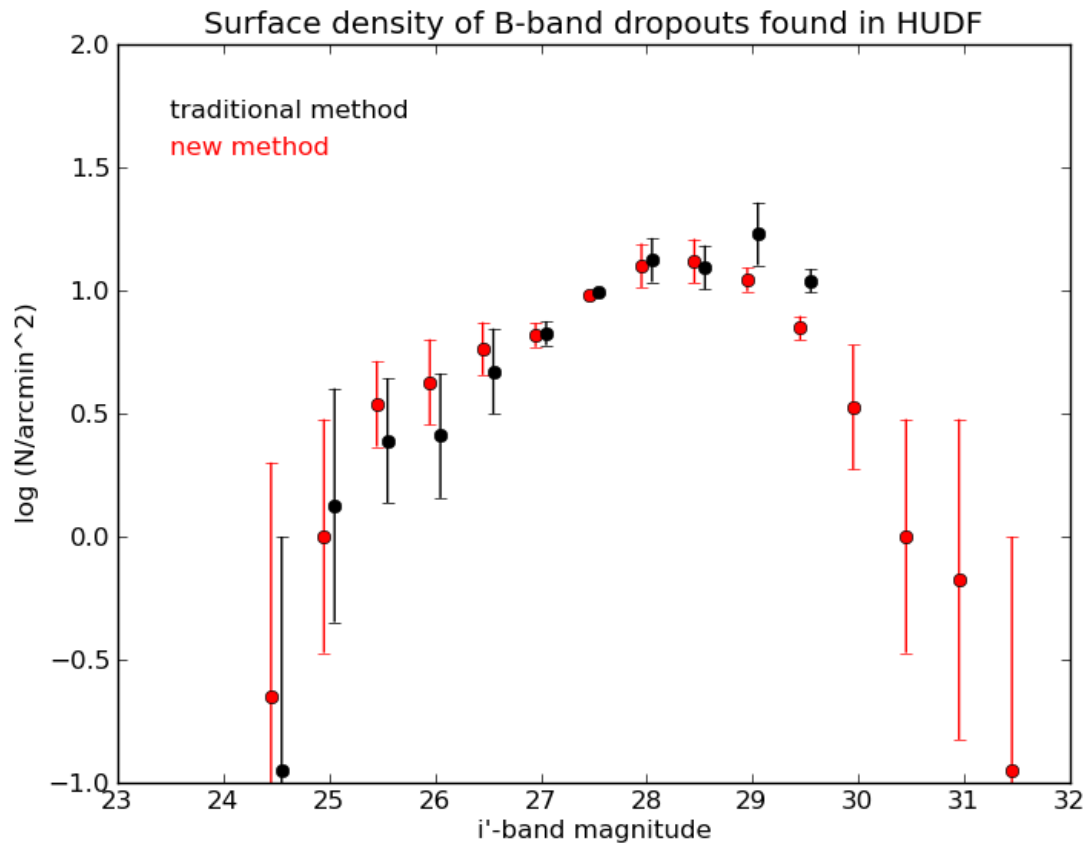
Figure 5.4: Surface density estimates as a function of i′-band magnitude: from the traditional dropout method (black dots), and the from the new method (red dots). The two sets results are offset from each other by 0.1 magnitude for clarity.

6 Discussion

How do the selections made using the new method compare with those made by the traditional method? Could there be galaxies at $z \sim 4 - 5$ that were missed by the traditional dropout detection method, but identified by the new method? Figure 5.2 shows that the new method is detecting a greater variety of high-z galaxies (blue dots) than the traditional method (red dots). It is possible that these detections could include new galaxies to add to the sample of galaxies that were available to reionize the universe, or that they could be contaminants. When the colors of these RSF-selected galaxies were compared to colors of dropouts found by R. Bouwens [Bouwens, et al., 2007], they were in agreement. Also, our method found no galaxies with V-i′ redder than that of Bouwens' $z \sim 6$ galaxies, which is what we expected, given our data set. When matched with galaxies having photometrically determined z's from the Coe catalog [Coe et al., 2006], we found that ∼90% of the sources detected by the new method are at the correct redshift. This means only ∼10% of our RSF-selected set of galaxies have $z < 4$, which is low contamination. Also, only <0.1% of our matched set were missed by the new method. The goal of this program was to find high-z galaxies, and to add to the existing set of these by finding new ones. The low contamination of the sample and the extremely tiny number of missed detections are key in attaining this goal. All of these results are evidence that the new method correctly identifies galaxies at $z \sim 4 - 5$.

Figure 5.4 shows that our calculated surface densities of high-z galaxies are comparable to those resulting from the traditional method [Bouwens, et al., 2007], and slightly more robust at brighter magnitudes. This comparison is an important check: the new method identifies as many

SExtractor-detected sources as the traditional method.

One important future step is the is use of the redshift factor image *without* the segmentation maps that select galaxies with a high signal-to-noise ratio. The redshift factor image is a collection of pixels that could hold light from sources at $z \sim 4$ missed by SExtractor due to their low flux levels. If the light from these pixels were summed, we could account for this missing $z \sim 4$ light. When this technique is eventually extended to probe $z \sim 6-20$ (reionization era), this summed light could be compared to the amount of light required to reionize the universe in order to determine whether early galaxies are candidates for producing this light.

7 Conclusion

Since the 1990's, the dropout method has been successfully employed on deep, multiwavelength, optical images of the Hubble Ultra Deep Field (HUDF) to identify Lyman break galaxies at $z \sim 4 - 5$ that have high signal to noise ratios. Here we developed and tested a new method for identifying Lyman break galaxies at $z \sim 4 - 5$ whose signal is on the order of the noise level in the HUDF images. We detected sources in the images using SExtractor on a detection image made from combining the V- and i'-band images. A redshift factor image was created by combining gaussian-smoothed images in a way that the redshift factor image would be $\sim$1 for galaxies at $z \sim 4 - 5$ and $\sim$0 otherwise. When the redshift factor image is used in combination with a segmentation map of source locations, $z \sim 4 - 5$ galaxies can, in theory, be identified. We found that:

- 749 out of 5370 sources had a RSF value that selected them as having $z \sim 4 - 5$

- these 749 had colors that placed them at $z \sim 4 - 5.5$

- when matched with a catalog of sources containing photometric redshifts [Coe et al., 2006], 4811 of the sources also were classified as being in $0 < z < 5.5$

- 90% of these were identified as being at the correct z by the new method, and <0.1% of galaxies at $z \sim 4$ were missed by the new method.

- surface densities of RSF-selected high-z sources calculated by us were similar to surfaces densities from earlier work by R. Bouwens [Bouwens, et al., 2007].

We conclude that the new method produces a robust sample of high-z galaxies. In the future, the redshift factor could be used to select individual pixels that have light from high redshift. This could provide a measurement of light from high-z galaxies that are not detected as individual sources. When this technique is eventually extended to probe $z \sim 6 - 20$ (reionization era), this summed light could be compared to the amount of light required to reionize the universe in order to determine whether early galaxies are candidates for producing this light.

Bibliography

[Barkana & Loeb, 2001] Barkana, R., Loeb, A. *In the Beginning: The First Sources of Light and the Reionization of the Universe*. 2001, Phys. Rept., **349**, 125-238

[Beckwith, 2006] Beckwith et al. *The Hubble Ultra Deep Field*. 2006, AJ, **132**, 1729

[Benitez, 2000] Benitez, N. *Bayesian photometric redshift estimation*. 2000, ApJ, **536**, 571-583

[Bertin & Arnouts, 1996] Bertin, E. & Arnouts, S. *SExtractor: Software for source extraction*. 1996, A&AS, **117**, 393

[Bouwens, et al., 2007] Bouwens, R., et al. *UV Luminosity Functions at z 4, 5, and 6 from the Hubble Ultra Deep Field and Other Deep Hubble Space Telescope ACS Fields: Evolution and Star Formation History*. 2007, ApJ, **670**, Issue 2, 928-958.

[Coe et al., 2006] Coe, D., Benítez, N., Sánchez, S. F., Jee, M., Bouwens, R., & Ford, H. *Galaxies in the Hubble Ultra Deep Field: I. Detection, Multiband Photometry, Photometric Redshifts, and Morphology*. 2006, AJ, **132**, 926

[Giavalisco, 2002] Giavalisco, M. *Lyman-Break Galaxies*. 2002, Ann. Rev. Astron. Astrophys., **40**, 579-641

[Guhathakurta, 1990] Guhathakurta. P. *A Redshift Limit for the Faint Blue Galaxy Population from Deep U Band Imaging*. 1990, ApJ, **357**, L9-L12

[Gunn & Peterson, 1965] Gunn, J.E., Peterson, B.A. *On the density of neutral hydrogen in inter-galactic space.* 1965, ApJ, **142**, 1633-1641

[Madau, 1995] Madau, Piero. *Radiative Transfer in a Clumpy Universe: The Colors of High-redshift Galaxies.* 1995, ApJ, **441**, 18-27

[Oke & Gunn, 1983] Oke, J.B. & Gunn, J.E. *Secondary Standard Stars for Absolute Spectrophotometry.* 1983, ApJ, **266**, 713-717

[Penzias and Wilson, 1965] Penzias, A. A. & Wilson, R. W. *A measurement of excess antenna temperature at 4080 Mc/s.* 1965, AJ, **142**, 419-421

[Robertson, et al., 2010] Robertson, B. E., Ellis, R. S., Dunlop, J. S., McLure, R. J. and Stark, D. P. *Early star-forming galaxies and the reionization of the Universe.* 2010, Nature, **468**, 55