North Carolina Journal of
Mathematics and Statistics
Volume 00, Pages 000–000 (Xxxx XX, XXXX)
S 0000-0000(XX)0000-0

Visual Perceptions about Gaussian Noise Filtering on Grayscale Low-light Photography

Aidan J. Draper and Laura L. Taylor

ABSTRACT. Statisticians, as well as machine learning and computer vision experts, have been studying image enhancement through denoising different domains of photography, such as textual documentation, tomography, astronomical, and low-light photography. With the surge of interest in machine- and deep-learning, many in the computer vision field feel that current approaches for effective image denoising are moving away from statistical inference methods and, instead, moving into these subfields of artificial intelligence. However, this paper sheds some light on current applications that show how statistical inference-based methods and frameworks that rely on conditional probability and Bayes theorem are still prevalent today. We reconstruct a few inferential kernel filters in the R and python languages and compare their effectiveness in denoising RAW images. In doing so, we also surveyed Elon students about their opinion of a single filtered photo using the various methods. Many scientists believe that noise filters cause blurring and image quality loss so we investigated whether or not people felt as though improved denoising causes any quality loss as compared to their original images. Individuals assigned scores indicating the image quality of a denoised photo compared to its true no-noise image on a 1 to 10 scale. Survey scores for the various methods are compared to benchmark tests, such as peak signal-to-noise ratio, mean squared error, and r-squared, to determine if there were any correlations between visual scores and the benchmark results. (NEED TO EDIT - THIS IS STILL THE ORIGINAL)

1. Introduction

The world is full of many naturally-occurring signals. Because there are so many, it becomes rather hard to isolate them at times. Photographers try to isolate light signals during exposure and many other domains of science also study signals that can be even harder to single out. Even with the immense collection of advanced electronics found in today's average digital camera, captured light signals are still prone to random noise from sources such as the camera's sensor or circuitry and as a response to qualities of the environment, like temperature. Of the environments in which photographers love to shoot, none resist noise as poorly as low-light environments do. As popular of a time as dusk is to capture, it can also be the most detrimental to image quality. This is what makes the field of signal denoising so compelling. Many researchers have spent much of their lifetime investigating methods for properly denoising signals. In doing so, standardized formulas have been developed for validating the effectiveness of these methods. However, past studies have reported that the benchmark scores typically associated with these methods are uncorrelated

Received by the editors December 14, 2018.
2010 *Mathematics Subject Classification.* 11A11; 00B12.
Key words and phrases. Image denoising; Image quality loss; Statistics in signal processing; Computer vision methods; Box-filtering methods; R Shiny applications; Python OpenCV package; Low-light photography; Gaussian grayscale noise.

with unbiased observers' opinions of denoised-image quality (NEED TO GRAB THESE REFERENCES). Those past studies were a motivation for continued research in the investigation of image quality in this paper.

In studying image quality, it was apparent that there is a void in computer vision literature surrounding visual perception of image quality in denoising methods. Additionally, this study differs from past work for three main reasons. First, it focuses on the subdomain of low-light photography due to current interests surrounding this setting in social media. This meant creating a unique low-light image dataset to suit the study. Second, quality of filtered grayscale images was investigated and visual perceptions of college students were collected. This was for convenience and to offer a different perspective than previous studies have had. Third, actual noise was captured rather than simulating the disruption of random pixels on a noiseless image. This adds some complications later, but also, provides real world examples of the filter methods' abilities. Lastly, a proprietary image denoising method in Adobe's creative cloud was included. This is by far the most accessible method to photographers and has also received a fair amount of criticism in the photography community, which made it an interesting addition to this study.

In this paper, five filter methods' performances on a single noisy photo were evaluated. After implementing these filtering methods on a single image from our low-light dataset, the benchmark scores of the filtered images were compared. Then, a survey collected undergraduate college students' perceptions of the filtered images' quality in relation to a noiseless image. Responses were collected and a one-way ANOVA test and ANCOVA test were performed on the visual perception scores about the image quality of various filtered images in relation to their desired non-noisy state.

This paper is organized in the following way. In section 2, we describe the background surrounding the subdomain of signal denoising in computer vision and, in doing so, share the methods performed on our test image. Section 3 elaborates on our methodology and experimental design, which includes the benchmark dataset we created, the R Shiny App survey that was built, and the process in which the experiment was performed. Section 4 shares the results of the benchmark tests on the single image as well as survey participants perception of the filtered photos. Finally, in section 5 we discuss the conclusions that we draw from the study.

1.1. Background

Computer vision is believed to be driven by two main pursuits for knowledge. From one perspective, scientists look to model human vision processes. Interest surrounds mimicking common human ability and understanding how human perception and comprehension occurs. On the other end of spectrum, scientists look to improve autonomy in machines and perform advance tasks, such as identifying objects or understanding dynamically-changing scenes, unrelated to understanding how human vision works (Zhang, 2015). These philosophies inevitably overlap at times, but these motivators are what drive the study of vision in computer science. The field is said to have emerged partly as a result of Larry Robert's thesis at MIT, where he introduced the concept of extracting 3-dimensional shapes from 2-dimensional images using "line drawing" to retrieve edge information (Roberts, 1963). Many scientists would follow in his footsteps in studying a subdomain of computer vision that is today known as edge detection. Self-driving vehicles are some of the newest technology that require advanced methods for scene understanding that mimic human perception, while also performing many other processes that go beyond our human capability, like tracking distance from other vehicles.

The study of noise removal would be motivated by the latter pursuit. Signal denoising experts look to filter noise to improve pre-processing for machine autonomy and a broad scope of other science processes. Image denoising has been used to cleanup tomographic photos for electron identification (Fernandez, 2009). It has also been used to process and cleanup many other signals, including electronic hisses, magnetized particles from magnetic film, tendrils and candlesticks from stock market data, and grain from satellite images. The specific study of digital image denoising emerged following the invention of the first charge-couple device camera in 1975. (NEED TO ADD MORE ON THIS HERE)

In the field of signal processing, there are a few different types of noise that haunt images. What makes statistics useful is that most types of noise follow the same probability density functions of some common random variables. The most common found in low-light photography are Gaussian noise and Poisson noise, which follow the distribution of their respective random variables. (NEED MORE ON DISTRIBUTIONS)

There are many heuristic methods for noise removal, but none have been able to fully denoise a scene to this date. Competition in the field emerged for the title of ”best-approach”. Noise reduction is computationally demanding though so methods are judged by both, their effectiveness and running-time. The need to evaluate algorithms with slightly-differing, approximate results led to the development of common scores for judging denoised photos. The most common include mean squared error (MSE), r-squared, peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and run time of course. Their formulas are described below. For reference ”true state” and ”filtered state” describe the respective pixel matrices of the imported two images.

- Mean squared error = $\frac{|\text{filtered state} - \text{true state}|^2}{\text{N}_{\text{True State}}}$
- R-squared = $\frac{1 - \left(\text{true state} - \text{filtered state}\right)^2}{\left(\text{true state} - \mu_{\text{true state}}\right)^2}$
- Peak signal-to-noise ratio = $20 * log_{10}\left(\frac{\text{R-squared}}{\text{MSE}}\right)$
- Structural similarity = $\frac{\left(2\mu_{\text{true state}}\mu_{\text{filtered state}} + c_1\right) * \left(2\sigma_{\text{true state, filtered state}} + c_2\right)}{\left(\mu^2_{\text{true state}} + \mu^2_{\text{filtered state}} + c_1\right) * \left(\sigma^2_{\text{true state}} + \sigma^2_{\text{filtered state}} + c_2\right)}$

(NEED TO ADD MORE ON THIS HERE)

In this paper, we explore the effectiveness of a few of the most common noise removal methods shared by all sciences, but for the purpose of cleaning up Gaussian noise grain from low-light images.

1.2. Filter Methods

There are many strategies implored in noise removal. Motwani et al. dissect and classify most methods currently used in the field in a tree graph (Mukesh C. Motwani and Frederick C. Harris, 2004). They classify every algorithm into two overarching categories: Spatial Domain and Transform Domain. Spatial domains implement box filters, while transform domains are slightly more complex in their classification, but can be generalized to relying on a standard basis function that differs from the traditional box filter. As mentioned earlier, the study was conducted using five of the most common, and arguably simplest in the field, image denoising methods. They include the bilateral filter, the non-local means filter, the three-by-three mean filter, and two levels: 50% and 100%, of the Adobe Lightroom CC denoise filter. Each method will be dissected in this section.

1.2.1. Three-by-three Mean Filter

The mean filter is a standard approach to noise filtering. It implements a box, sometimes called spatial, filter of size z by z. This box filter is a smaller matrix that traverses the larger image matrix averaging just the center value of every box along the way. This box matrix is typically denoted as x with a size of n, or $z * z$. The method is classified as a spatial domain linear filter known to be used to specifically target a decrease in mean squared error. The mean filter has received criticism for destroying edges, destroying fine details, and blurring lines in images (Mukesh C. Motwani and Frederick C. Harris, 2004). It is expressed symbolically as follows:

$$I^{filtered}(x) = \frac{1}{n} \sum_{x_i \in \Omega} x_i$$

where Ω represents the entire image and $I^{filtered}$ is the resultant filtered image. This experiment implores the ability of a three by three filter, which is a typical size for this method that causes less blurring than some of the larger spatial filter sizes.

1.2.2. Non-local Means Filter

One of the first public introductions of the non-local means filter was during an IEEE conference in 2005 by Baudes, Coll and Maurel (A. Baudes and Morel, 2005). The algorithm relates to many linear filtering methods, but it calculates the weighted probability impact of each pixel for averaging the center value based on the similarity of neighboring pixel scores within the box instead of using a standard approach for probability weights, such as a Gaussian probabilities or equal weights. It is expressed symbolically as:

$$I^{filtered}(p) = \frac{1}{C(p)} \int_{\Omega} v(q) f(p,q) dq.$$

For this study, the OpenCV Non-local means filter was specifically implemented rather than the formula because OpenCV's method has been optimized to decrease runtime as much as possible.

1.2.3. Bilateral Filter

The Bilateral Filter is a far more advanced filter that takes into account 3-dimensional space by also considering whether or not the next pixel would be unlikely to see based on a distribution of the previously viewed pixels. This is expressed by the g_s in the formula by calculating residual error. Symbolically, this formula is:

$$I^{filtered}(x) = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) f_r(||I(x_i) - I(x)||) g_s(||x_i - x||)$$

Again, the OpenCV bilateral filter was specifically implemented rather than the traditional formula because OpenCV's method has been optimized for runtime.

1.2.4. Adobe Lightroom Denoise Filter

At the time of this study, there was no public information shared about Adobe's methods. It appears they have sacrificed some performance to be scalable on larger images and to output at an acceptable runtime to users. At 100%, there method appears to blur edges at an extreme rate. Photos look practically washed and fine details are nonexistent.

2. Methodology

Noise is considered to be a quality that negatively impacts the perception of an image. One of the most typically seen random variable distributions of noise is Gaussian noise, which is an unwanted channel that is captured during the camera's acquisition of the desired light wavelengths. Many have explored methodology for replacing the unwanted pixels with computed pixels that replicate what the desired wavelengths may have shown using a variety of mathematical approaches for approximation and estimation. The common approach for testing these methods involves mimicking a noisy image and then, taking a true image with no noise so that they can compare how well their computed pixels replicate the photo that would be originally desired. They use the standard benchmark scores, PSNR, MSE, and R-squared, as we previously mentioned, to score their results.

This study models that process of analyzing methods' abilities to approximate missed pixels of the true signal, but goes a step further by surveying college undergraduates to investigate whether they believe image quality has in fact improved. The exact process implemented will be expanded on below.

2.1. Protocol

The first step was to compile an original dataset of true and noisy images to test. Although there are image datasets that exist, such as the RENOIR and Darmstadt, none look to specifically model low-light photography, which was of interest in this study. The camera model used was a full-frame Canon 6D Mark I released back in 2007. It had a USM 17-40mm f/4L Canon lens attached to it. Two photos were taken at each location with greatly varying settings. First, the original photo would be shot with an ISO of around 200-600 depending on the scene. The aperture was set to the lowest setting in order to let the most light in. The shutter of the camera would be left open for 1 to 5 seconds in order to properly expose the photo at such a low ISO setting. The second photo was shot using the same aperture. However, the ISO setting was increased to just below the camera's maximum light sensitivity settings, which was around 20000 or 25600 ISO, for different scenes to dramatize the amount of noise captured in the noisy image. There are approximately 20 million pixels in each image so noise will not be as apparent when presented regularly.

FIGURE 2.1. A true state image (left) and a noisy image (right) taken with different camera settings.

After photos has been captured, they were initially brought into Lightroom where the same grayscale filter was applied to every image. Images were then exported out of their CR2 Canon RAW image formats and into the best possible resolution JPEG format. Python filter scripts for the Bilateral filter, Nonlocal Means filter and three by three Mean filter were applied to the noisy image

FIGURE 2.2. An example of an image with a blown up region.

of interest. The unfiltered noisy images were also brought back into Adobe Lightroom so that the Denoise filter could be applied at the 50 and 100% levels. The respective filtered photos were saved and loaded into R. An R script grabbed the true image and the filtered image and converted both into matrices. Then, PSNR, R-Squared and MSE were calculated for each filter variation for the same subject matter.

To capture student perceptions of the filtered images, a survey was built using a Shiny App in R. A specific frame was selected so that there was a control in subject matter presented to participants. The subject matter selected for the noisy, filtered and true state images can be seen in Figure 2.1 from earlier. The survey began by presenting respondents with a noiseless photo and then, asking them to rate three following photos of the same content, but with varying degrees of image quality. Once the initial training portion was complete, respondents were asked to rate a final photo about its image quality on a scale of 0-10 in comparison to its respective noiseless image. A score of zero represented an image that had far inferior quality as compared to the original. A score of five indicated equal image quality to the respective noiseless image. Lastly, a score of ten meant that the respondent felt that the photo had far superior image quality to the noiseless photo. The first training photo presented was the unfiltered noisy photo. The second training photo was a Nonlocal Means filtered noisy photo. The third training photo was the same photo as the noiseless photo. These three scores were meant to serve as an indication of an inability to identify noise, or image quality, in respondents while also taking into account that respondents will have different metrics for image quality, naturally.

Once the respondent had selected scores for those three training images, they proceeded to the images of interest. Again, the noiseless image was presented first and then, they were asked to compare a randomly-selected filtered photo to the identical noiseless photo. To emphasize differences in images, a blown up region of each image was included in the bottom right-hand corners. These blown up regions were 600 by 600 pixel grids from sections of the photo that included an edge. In Photoshop, these regions were doubled in size to approximately 1200 by 1200 pixels. An example of this can be seen above in Figure 2.2. After the participant rated the last image, scores were sent to a Google Sheet.

2.2. Survey

Introductory mathematics and statistics students attending Elon University were chosen as a sample population in order to expose younger students to undergraduate research. Additionally, it provided a convenient sampling frame for the time restrictions of this project. Under IRB approval, instructors from the previously mentioned courses were emailed and asked to share the survey link with the students in their sections. 89 responses were recieved, meaining that the survey had far more non-respondents than respondents. Images of the survey are included in Figure 2.3 below Figure 2.3. There were a total of nine slides in the survey, which are displayed in order in Figure 2.2. The survey included:

(1) Introduction slide - describes the survey, who to contact, and that response is optional,
(2) First instructions slide - informs users how to go about rating the following images,
(3) First true state image slide - a reference image to compare the following training images to,
(4) Training image one slide - an unfiltered noisy image with a horizontal score bar at the bottom,
(5) Training image two slide - a filtered noisy image with a horizontal score bar at the bottom,
(6) Training image three slide - the same true state image with a horizontal score bar at the bottom and a ”Submit Part One” button that leads to the second part of the survey,
(7) Second instructions slide - informs the user how to go about rating the following image,
(8) Second true state image slide - a reference image to compare the filtered image of interest to,
(9) Filtered image of interest slide - a randomly-selected filtered image (either Mean, Nonlocal, Bilateral, Abobe 50% or Adobe 100%), a horizontal score bar to rate the image, and a ”Submit Part Two” button to send the scores to a Google Sheet.

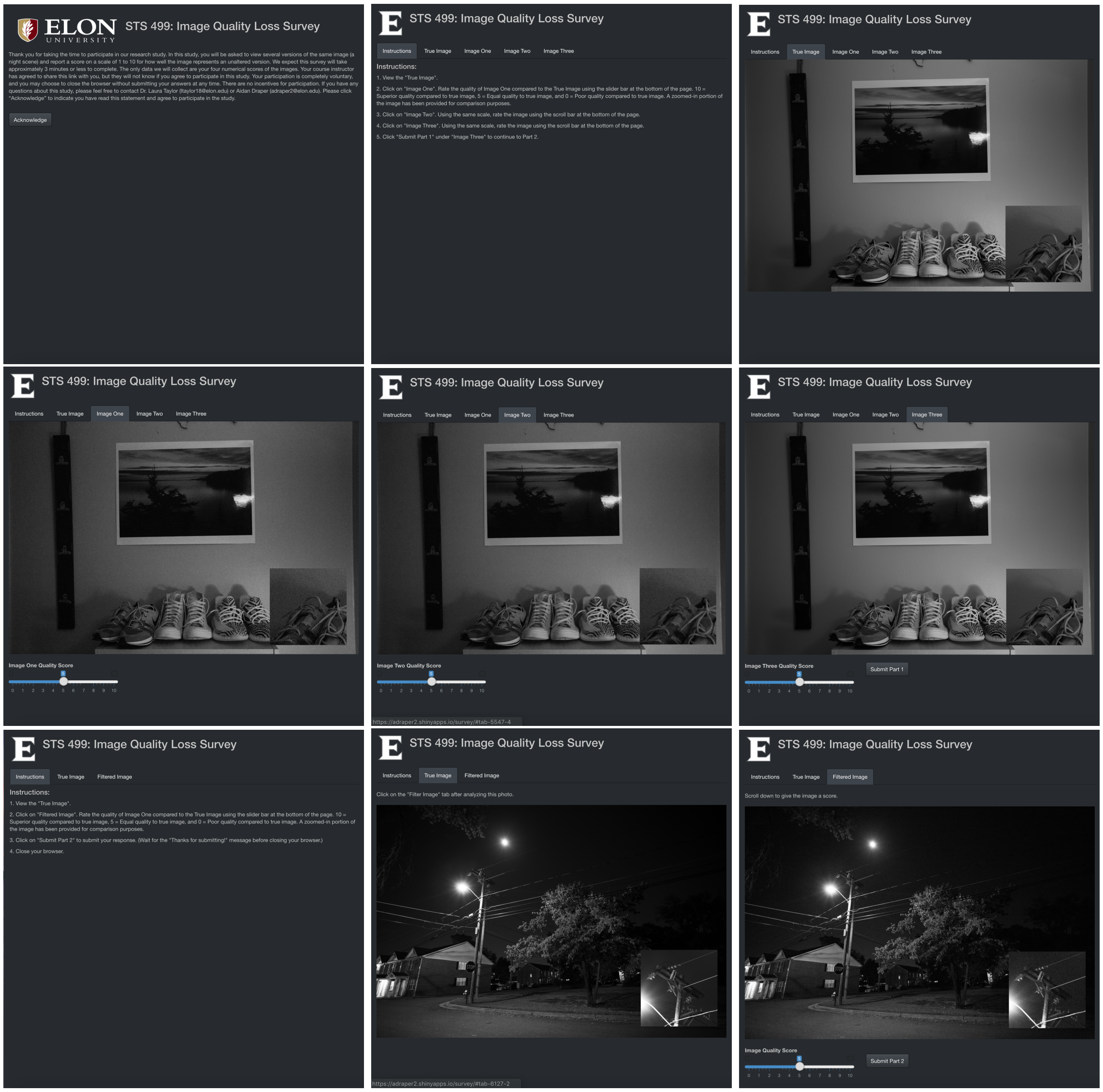

FIGURE 2.3. Slides from the Shiny App survey.

3. Experiment Results

The five filtered images with their blown up regions are listed next to the original noisy image that was filtered in Figure 3.1. Visually, the Adobe 50% and Bilateral filter produced the most compelling filtered images. There was the least amount of loss of fine-detail as well as edge preservation. Additionally, the images did not show any signs of wash, which is due to excessive blurring. Noise appeared less present in these images as well. The least compelling was the Non-local means filer, which did not preserve edges or fine details well and also, still has a degree of noticeable noise still present. The mean filter was the worst at removing noise, but detail and edges have been preserved due to the lack of blurring. The Adobe 100% has an awful amount of blurring even though noise is nonexistent in its resultant image.

In general, the results were somewhat expected. Benchmark results for the Mean, Non-local mean, and Bilateral methods followed similar results to other papers. All filters scored higher than

FIGURE 3.1. Filtered images in comparison to the original noisy image. Images are ordered from left to right and top to bottom as follows: Noisy, Mean, Bilateral, Nonlocal, Adobe 50, and Adobe 100. (THIS PROBABLY NEEDS TO BE PRESENTED DIFFERENT - HARD TO DETECT ANY DIFFERENCES)

the unfiltered noisy image. In Table 3.1, we can see the exact scores received. Peak signal-to-noise ratio improves as the number grows. Mean squared error improves as the error decreases between true image and filtered image. R-squared improves when the true image and filtered image share more and more similar pixels between one another in (x, y) positions. It is bounded by 0 and 1. The Adobe 50% filter had the least improved scores, which was the most unexpected result. The photo in Figure 3.1 had far less blurring than a few of the other filters. Blurring did not seem to get punished by the benchmark scores though. The Adobe 100% filter had the second most improved results, but was by far the most blurred image of the five filters. The best performing filter was OpenCV's bilateral filter, which also was one of the more visually-compelling filtered images. The non-local means filter had one of the worst appearances because it seemed to be left in two states.

It was not quite denoised and had lost a decent amount of its original sharpness, which is a rather disappointing resultant image quality for what is considered to be a standard filter in the field. However, it scored fairly well in comparison to the other methods.

TABLE 3.1. Benchmark Results for the Street Lamp Image.

	Unfiltered	Mean	Bilateral	Nonlocal	Adobe 50%	Adobe 100%
PSNR	41.7691	46.8833	47.8103	46.9004	46.3890	47.5747
MSE	530.4268	294.3851	264.8607	293.8056	311.6243	271.8604
R-squared	0.4426	0.6732	0.7217	0.6913	0.6732	0.7143
SSIM	0	0	0	0	0	
Runtime	0	0	0	0	0	

Figure 3.2 depicts the distribution of image quality scores for all methods. Interestingly, more than one person rated methods as 0, implying terrible quality in comparison to the noiseless image, and as 10, implying far superior image quality in comparison to the noiseless image. Most respondents scored images as either 3 or 4, which is what was expected. The slightly right-skewed distribution was unexpected though.

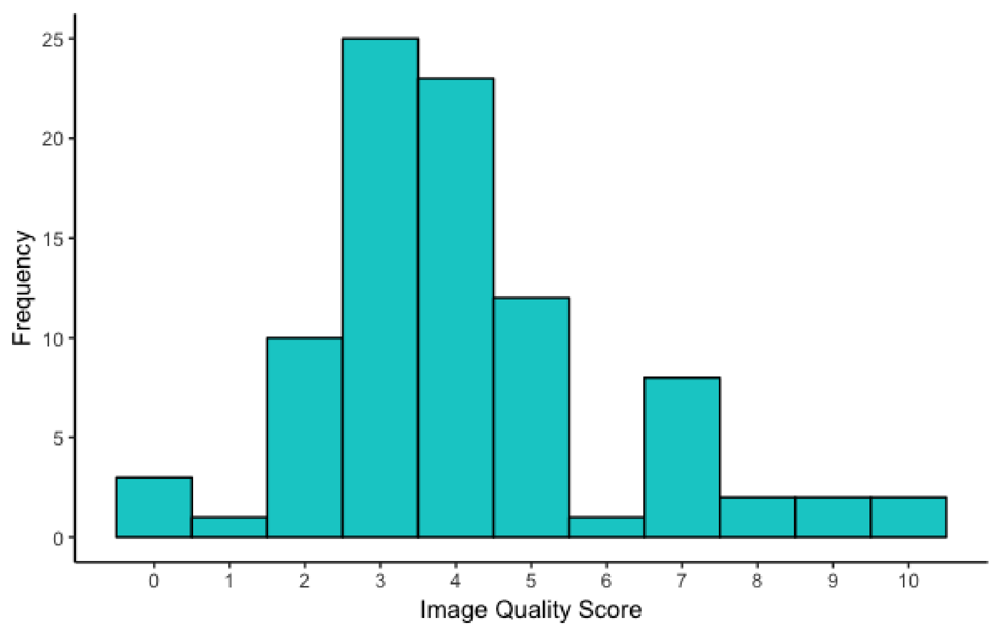

FIGURE 3.2. A histogram of the participants' image quality perception scores.

When distributions are broken down by methods in Figure 3.3, individual distributions of respondent image quality scores can be compared. For the most part, group means remain fairly similar. The bilateral filter surprisingly had the smallest median of approximately 3, while all other methods had a median of 4. The mean of the Adobe 100% method was the largest at roughly 5. The smallest mean was the Three-by-three mean box filter. That being said, almost every method's mean is being skewed by either a large range of values or an outlier. Adobe 100% received almost every score and its range captures the entire scoring range offered. The overlap in interquartile ranges suggests a non-significant p-value from the ANOVA test.

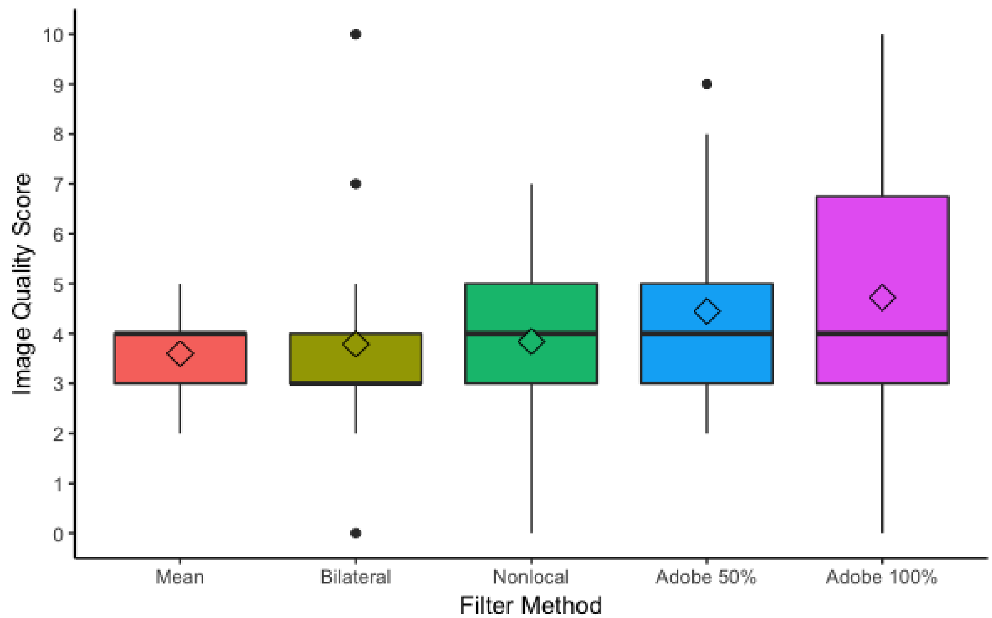

FIGURE 3.3. Boxplots of the participants' image quality perception score distributions by filtering method.

As expected, the ANOVA test yielded insignificant results. Based on a p-value of 0.432, there is very insufficient evidence of a population mean difference in respondents' image quality scores based on the noise filtered method implemented. Much more of the residual error is still found within individual methods. This is likely due to the great variation in scores received from respondents for individual methods.

TABLE 3.2. An ANOVA table for the image quality perception scores.

	DF	SSE	MSE	F-score	P-value
Treatment	4	15.9	3.985	0.964	0.432
Error	84	347.3	4.135		
Total	88	363.2			

Almost all assumptions were passed before performing this ANOVA. The Q-Q plot, as well as the scatterplot of the residuals against the fitted values, are shared in Figure 3.4. Assumptions about additive qualities of the model and fixed population means are passed due to no indication of curvature in the scatterplot. Assumptions about residuals, including shared standard deviations, independence, and averaging to zero appear to also pass as there is no indication of a cornucopia or pattern in the residuals against the fitted values. The Q-Q plot is slightly concerning because it shows some curvature in its tails due to outliers. The residuals are not quite normal, which will be a limitation in this study. The neither variance nor sample sizes are equal, which would lead to less confidence in any significance level.

When analyzing the data, it became apparent that there may be a relationship in training scores and the final image scores received. For this reason, an ANCOVA test was conducted to determine if there is any effect when the average training score of the initial three images increases between respondents. Results from this test can be seen in Figure 3.5. (I THINK WE SHOULD

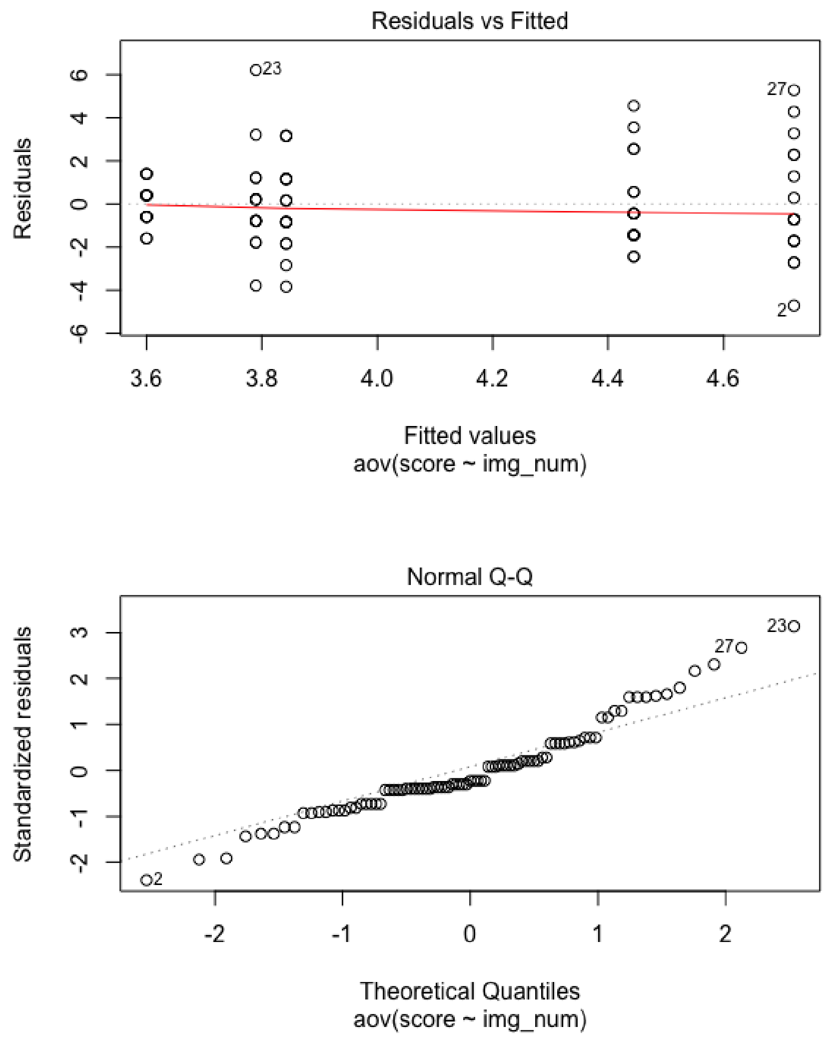

FIGURE 3.4. A Q-Q plot and scatterplot of the residuals for the ANOVA test.

TALK ABOUT THIS FIGURE AGAIN, I AM STILL SOMEWHAT UN-CONFIDENT TALKING ABOUT WHAT AN ANCOVA TEST IS IN THIS PAPER).

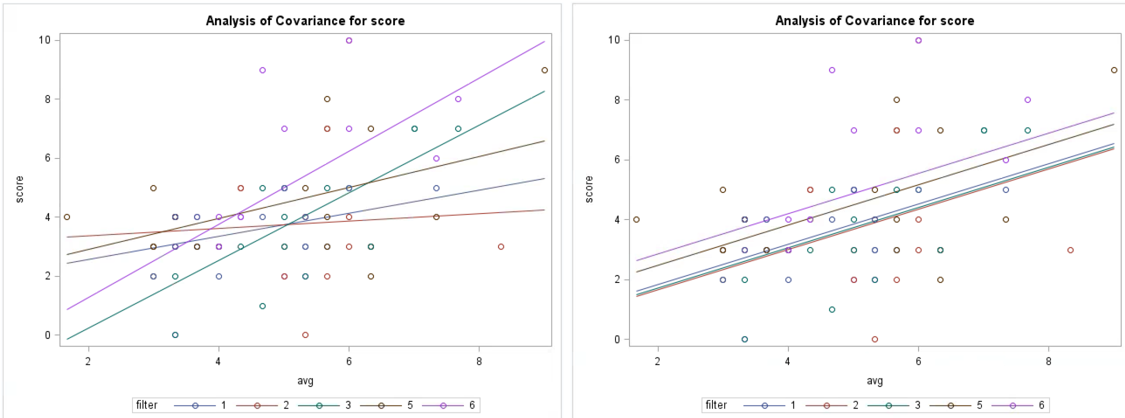

FIGURE 3.5. Figures from the ANCOVA test. (DID NOT TALK ABOUT YET)

4. Discussion

There was little variation in types of methods used. Many of these algorithms are linear filters, which are considered to be somewhat primitive in comparison to some Transform Domain methods. This led to benchmark scores being somewhat indifferent. There were noticeable differences, but some may argue that they would be insufficient to warrant the use of one method over another based solely on score. Visually, the filter methods led to quite different results though.

Results were insignificant from the ANOVA test when comparing sample populations between filtering methods. This result is interesting though because it signifies that, based on the experimental design, respondents could not detect a significant difference between methods. This could be due to the size of the image used. Perhaps a larger blown up region or smaller original image may make individual pixel differences more apparent. Additionally, benchmark scores varied, but were rather close between methods as previously mentioned. There is some argument that greater differing benchmark results may lead to greater visual perception differences.

Based on median and means alone, it was surprising to see how much higher the scores for the Adobe 100% filter were compared to the Bilateral filter method. The averages for image quality score responses from participants was somewhat inverse to the benchmark scores. The best performing filter according to the benchmark scores was the worst performing filter in the eyes of survey respondents. Additionally, there was not trend in benchmark scores and visual perceptions of image quality overall. The two were somewhat unrelated, which gives some indication to either, an unrepresentative sample, a sample that does not properly understand image quality, or benchmark scores that do not properly score methods. Although it would be desirable to declare that scoring methods are not currently fitting, it is more likely that the sample population is not representative. A better sampling frame of individuals, who had a strong grasp of image quality, would be needed to draw a conclusion about the ability of current benchmark scores to properly evaluate methods.

One of the most surprising results was the variance of the Adobe 100% filtering method, which was expected to receive fairly negative results in general due to extreme blurriness. It appears that many of these minimum and maximum image quality scores also correlated to minimum and maximum scores recorded from the training set. Overall, the training set scores had an average of approximately 7.2, implying that, on average, the respondents felt that the three images had superior quality to the noiseless image. This is very unexpected considering that the third image was the same exact image being compared, which should warrant a image quality score of 5. The expected mean would be in the range of 3 to 4. For this reason, there is indication that respondents felt very strongly about filtered images or, more likely, participants did not fully read or understand the directions about the image quality scoring scale. If a 10 indicated equal image quality to the noiseless image, that may explain the odd amount of 6s, 7s, and 8s the filtered images received. The ANCOVA test did a good job of quantifying any possible visible difference between image quality scores for the filter methods that the mean training score may have accounted for. (NEED TO ADD DISCUSSION ABOUT ANCOVA HERE)

5. Conclusion and Future Work

The results of the ANOVA test led to no indication of a significant difference between sample means of methods implemented. On further analysis, training scores appeared to be useful indicators of bias in respondents. Testing for differences in training scores before final image scores led to some interesting explanation into the odd variances received for some methods. (WILL ADD MORE ABOUT ANCOVA IN FINAL VERSION)

There were complications in pursuing this study. The dataset appears to be ridden with bad data points most likely due to scale misunderstandings. This occurrence led to largely differing variances between sample populations of filter methods and because of that, an assumption of the ANOVA test was violated. In performing future tests on visual perceptions, either clearer instructions or an easier to interpret scale should be implored. In addition, greater variation in methods, such as comparing Spatial Domain to Transform Domain methods may lead to a greater variation in resultant image quality. Researches have noted that many Transform Domain algorithms are far superior in performance than many of the linear Spatial Domain methods used, including the mean and non-local mean methods implemented in this experiment. Time restrictions limited the application of some more advanced methods, including a Markov Random Field, which, if implemented, would most likely suffice in introducing more impressive noise filtering to the study.

This study was also performed on the notion that respondents should not be made aware of what noise was while scoring images. Giving an indication as to what to watch out for in images, including edge blurring and loss of sharpness, may lead to a better understanding of image quality prior to scoring. Who is to say whether introductory-level mathematics and statistics students have an opinion about image quality that is not biased in comparison to the entire population of the college students. Also, college students are probably not the best population to ask about image quality. A more interesting sampling frame would be photographers in the United States. They would have a deeper understanding about desirable image quality. If more time was permitted, a greater size sample population may also lead to different results. In general, more emphasis should be made on retrieving accurate and representational perceptions about desirable image quality.

References

A. Baudes, B. C. and Morel, J. (2005). A non-local algorithm for image denoising. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Fernandez, J.-J. (2009). Tomobflow: feature-preserving noise filtering for electron tomography. *BMC Bioinformatics*, 10:178.

Mukesh C. Motwani, Mukesh C. Gadiya, R. C. M. and Frederick C. Harris, J. (2004). Survey of image denoising techniques. *Proceedings of GSPx 2004*.

Roberts, L. (1963). Machine perception of solids. *Massachusetts Institute of Technology*.
Zhang, Z. (2015). Image noise: Detection, measurement, and removal techniques. pages 1–10.

(Aidan J. Draper) DEPARTMENT OF MATHEMATICS AND STATISTICS, ELON UNIVERSITY, ELON, NC 27244, US
E-mail address, Corresponding author: adraper2@elon.edu
URL: http://www.aidandraper.com/

(Laura L. Taylor) DEPARTMENT OF MATHEMATICS AND STATISTICS, ELON UNIVERSITY, ELON, NC 27244, US
E-mail address: ltaylor18@elon.edu