EMBRY-RIDDLE
Aeronautical University
SCHOLARLY COMMONS

Doctoral Dissertations and Master's Theses

Fall 2023

Explorations in Monocular Distance And Ranging (MODAR) Techniques for Real-World Applications

Devon Vail
vaild@my.erau.edu

Follow this and additional works at: https://commons.erau.edu/edt

 Part of the Electro-Mechanical Systems Commons

Scholarly Commons Citation

Vail, Devon, "Explorations in Monocular Distance And Ranging (MODAR) Techniques for Real-World Applications" (2023). *Doctoral Dissertations and Master's Theses*. 780.
https://commons.erau.edu/edt/780

This Thesis - Open Access is brought to you for free and open access by Scholarly Commons. It has been accepted for inclusion in Doctoral Dissertations and Master's Theses by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.

Explorations in Monocular Distance And Ranging (MODAR) Techniques for Real-World Applications

A THESIS

Presented to the Department of Mechanical Engineering

Embry-Riddle Aeronautical University, Daytona Beach

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Mechanical Systems

Committee Members:

Dr. Christopher Hockley – Advisor

Dr. Eric Coyle

Dr. Brian Butka

By Devon A. Vail

B.S., 2021, Embry-Riddle Aeronautical University, Daytona Beach

December 2023

`

Acknowledgements

I would like to acknowledge and thank everyone who has helped me through my graduate degree. I would have been unable to complete my degree without the support of my numerous friends, family members, and professors at Embry-Riddle Aeronautical University who helped me along the way.

First and foremost, I'd like to thank my advisor, Dr. Hockley, for inspiring me to research this topic all the way back in my undergraduate career. This idea has evolved quite a lot over the years and, under Dr. Hockley's guidance, it has become the focus of this thesis. Additionally, I'd like to thank Dr. Butka and Dr. Coyle for their help with this thesis. Dr. Coyle's mentorship through many projects and classes has helped refine my passion for robotics and helped me become the student I am today. I'd also like to thank Dr. Patrick Currier especially for continually advocating for my success as a student and finding ways for me to be involved in the research on-campus; without his support this thesis would not have happened. I'd also like to express my gratitude for the assistance of Mike Bakula for sharing his expertise of optics with me.

Furthermore, I'd like to thank my family for their ongoing support of my studies and their encouragement to keep me pushing through. I'd like to thank my girlfriend, Katelyn, as well for her constant support and encouragement. She kept me motivated throughout my college career and kept me going through this thesis.

To all my peers and friends, thank you so much for your words of encouragement and for letting me bounce ideas off you when I got stuck.

`

Table of Contents

`

`

Table of Figures

`

Abstract

In this work, an initial prototype of a monocular camera system capable of retrieving depth-from-focus using a liquid focus-tunable lens is constructed out of hobby-grade photography equipment. This concept has been explored previously in laboratory settings using specialized equipment; this work seeks to determine the feasibility of retrieving depth-from-focus using commercially available components. To achieve this, an iterative exploration of existing techniques was performed to verify their utility in the final ensemble of processes to retrieve depth from 2D images. Initially, blurry images were simulated by applying Gaussian blur to test images to verify the functionality of a Laplacian of Gaussian-based algorithm capable of determining image clarity, a sliding gantry was then constructed to move a camera through the environment and test the image clarity algorithm on real-world data as well as test methods to create a composite image of the most in-focus pixels from a focal stack of images collected while the camera was in motion. Following this, the depth retrieval algorithm was tested on a geared lens setup in which a gear-driven fixed focal length lens was attached to a camera and driven such that the distance between the lens and the imaging sensor in the camera was varied to change the optical power of the lens. This setup suffered from several limitations but provided significant insight into the fundamental principles governing depth-from-focus retrieval. Finally, 12mm, f/6, Liquid Lens Cx Series Fixed Focal Length Lens from Edmund Optics was attached to a Raspberry Pi Global Shutter camera to retrieve depth from an environment. This lens can vary its optical power by applying a voltage to the liquid lens which can be done automatically from a microcontroller at a high rate of speed. This operated with limited success and produced a very noisy depth map and point cloud of the environment. This work concludes with suggestions for future work to significantly improve the depth retrieval functionality of the liquid lens setup.

`

Introduction

Depth perception in machine vision is a core challenge within the field of robotics with many proposed solutions (Flyps, 2023). The ability to resolve the distance of an object to the origin of an imaging sensor is critical for many applications such as mobile robots navigating through cluttered environments, autonomous vehicle navigation, machines for material sorting and handling, and crowd monitoring to name a few. There have been numerous applications of depth cameras in the field of robotics and automation; some of the more prevalent examples include the Xbox Kinect sensor (Cruz, Lucio, & Luiz, 2012), which used depth mapping to track a person's movement and the Hazard Avoidance Cameras (HazCams) on the Mars 2020 Perseverance Rover (NASA, n.d.).

Depth Retrieval Sensors

Generally, depth retrieving sensors fall into one of two categories: active or passive sensors. As the name implies, active sensors actively transmit energy into the environment. When discussing the energy transmitted by active imaging sensors, most emit infrared light into the environment. Structured light cameras, time-of-flight cameras, and all forms of LiDAR are active sensors. Stereovision cameras and traditional monocular cameras are passive sensors.

Active Sensors

Depth cameras present a unique way to garner information from the environment not readily available from traditional monocular cameras; primarily, they produce depth maps of the environment. A depth map is typically a 2D grayscale image that is the same size as the red-green-blue (RGB) images they were created from. The grayscale intensity in the depth map image corresponds to the distance the object or surface the pixel represents is from the camera. Depth cameras generate depth maps in a variety of ways including projecting light in the form of infrared lasers into the environment to find the range to select points. Structured light cameras are another

`

method of capturing 3D data from an environment. These cameras use a specialized projector, whose spatial positioning relative to the camera is known, to illuminate a scene with a known pattern of light. In completely flat scenes, the camera will perceive a very similar pattern to the one projected. In nonplanar scenes, the pattern will deform, and the amount of deformation can be used to determine the shape of the surface causing the deformation (Geng, 2011). In addition to these cameras, time-of-flight cameras can also retrieve depth information from an environment by illuminating the environment with a pulse of light, typically an infrared light source, and measuring the time the reflected light takes to return to the sensor. This time can be translated into a distance to the surface the light reflected off. Emitting infrared light into the environment and measuring the time it takes to reflect back is also how Light Detection And Ranging (LiDAR) sensors work. However, LiDARs typically involve rotating optical components, such as mirrors, to rapidly transmit infrared light into the environment. This is not true of every LiDAR sensor; flash LiDARs are solid-state devices containing no moving pieces and function nearly identically to the time-of-flight cameras discussed previously (Li & Ibanez-Guzman, 2020).

Active sensors can suffer degradation in functionality in brightly lit and outdoor environments as the reflected return from their emitted pulse of light can be washed out in the ambient light. Additionally, multiple active sensors in an environment may destructively interfere with one another as the reflection of light emitted by one sensor may be detected by a different sensor thus causing incorrect depth values (FRAMOS, 2023). Adding to this, LiDAR sensors can be costly sensors with high initial costs for sensor acquisition and integration. Despite this, active sensors, like LiDARs, are preferable in scenarios where precision and processing speed is of paramount importance (METTATEC, 2023). Measuring distance with a laser offers high precision measurements and point clouds from LiDARs typically have fewer data points to process.

`

Passive Sensors

Passive sensors are characterized by the fact that they do not emit energy into the environment. One method of depth retrieval includes triangulating the same points in two different photos from two different image capturing sensors separated by a known baseline distance. This type of image capturing device is typically referred to as a stereovision camera (Dubey, 2020). Stereovision cameras and other passive sensors also suffer from setbacks primarily associated will lighting conditions in the environment. Similar to LiDAR and time-of-flight cameras, if the environment is saturated with light, a depth or stereovision camera will struggle to resolve the depth of the environment. Poorly lit environments will also illicit degraded performance in passive camera-based systems (Hesai Technology, 2023). However, camera systems are significantly more cost-effective and excel at supporting software for object classification when compared to LiDAR sensors (Vincent, 2023). Given their cost-effectiveness, it is a worthwhile endeavor to explore methods to maximize the utility of camera systems. While it might seem intuitive to only adopt use of stereovision cameras, they too suffer from some notable drawbacks; namely, the baseline

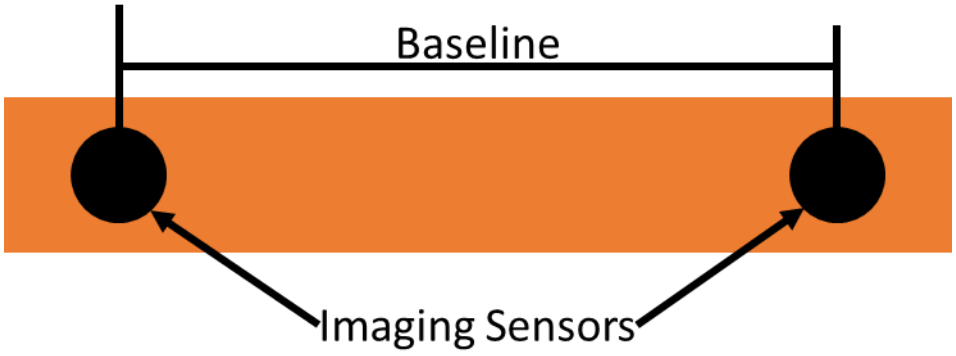


Figure 1: A graphical representation of the baseline of a stereovision camera

of the stereovision camera directly determines the range at which the camera can resolve depth. The baseline of a stereovision camera is defined as the distance between the imaging surfaces. Typically, the two imaging sensors in a stereovision camera are separated by only one degree of freedom, as seen in Figure 1. Doubling the baseline distance of the imaging sensors yields an

`

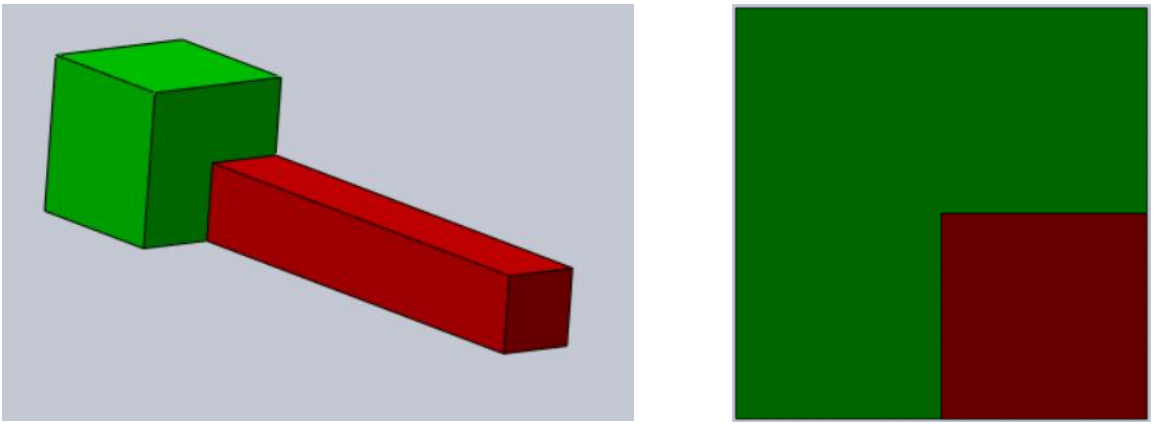
approximate 50% increase in the range of the camera (Dubey, 2020). Furthermore, stereovision cameras require more computational power as they are comprised of multiple imaging sensors.

Motivation of This Work

To avoid the pitfalls of a system with a large baseline and coordinating the data collection of two imaging sensors, it would be ideal to retrieve depth from a monocular camera to ensure a compact design and a single imaging sensor from which to process data. To this end, the remainder of this work will discuss depth retrieval from one imaging sensor using a primarily hardware driven approach referred to as depth from focus. Prior to discussing depth from focus in detail, it is necessary to some fundamental attributes of camera systems and the two-dimensional (2D) images that they capture.

Overview of Scale Ambiguity

One of the largest challenges associated with retrieving depth from a monocular imaging source is scale ambiguity. A rudimentary yet effective example of scale ambiguity can be seen in



a) b)

Figure 2: a) a green 20-inch cube attached to a red 10 x 10 x 70-inch rectangular prism. b) the same shape but from a different angle.

Figure 2. Figure 2a gives the viewer a clear view of the compound structure formed between the green and red rectangles whereas Figure 2b, which is an image of the same structure viewed from a different angle, Figure 2b offers no information regarding the length of the red rectangle. This may lead an observer to incorrectly assume that the face of the red rectangle and green rectangle

`

visible in Figure 2b are coplanar rather than separated by a great distance. The problem of scale ambiguity has been explored in a number of works *(Hong & Lim, 2018), (Fanani, Stürck, Barnada, & Mester, 2017), (Wang, Shen, & Chen, 2023)* with a large focus being put into scale retrieval for navigation and mapping algorithms such as visual odometry. Visual odometry in the use of vision sensors to locate obstacles and map them to real-world locations relative to the sensor. Once that has been done, the obstacles are tracked and, using these tracks, the vision sensors motion through the environment is computed. In such applications an accurate scale is critical to create accurate odometry estimates and avoid collisions.

While research has been conducted into resolving scale ambiguity, many methods require diligent cooperation from an operator or a known reference during the initialization phase of the algorithm. For example, Klein and Murray (2007) devised an algorithm for tracking and mapping movement in a small environment using a monocular camera. However, the algorithm required a user to initialize it using the press of a key and then to move the camera as smoothly as possible through a translational movement and then press a key again to end the initialization process. This kind of initialization leaves much to be desired for a product developed for end-users and purchasers of commercial products. In a manner similar to a user led initialization of the algorithm, Bleser et al. (2006) propose a camera pose estimation algorithm that requires the user to manually adjust the position of the camera until it is approximately aligned with an object in the scene of which a line model has been previously constructed. The authors refer to this method as a semi-automatic model-based approach due to the fact that the algorithm is not fully automatic, requiring the user to position the camera in an initial state that can view an object of which a 3D model has been created. From that 3D model, a 3D contour model is made which can be aligned with the 2D contours of the item in the monocular camera's field-of-view (FOV). As with the algorithm

`

requiring a user to perform specified actions at every initialization of the algorithm, producing a 3D contour model of an object and viewing that object every time one initialized the algorithm is unsuitable for end users of a commercial product and is generally unsuitable for most research cases as well.

As one can observe, the issue of scale ambiguity has generated a great deal of research. Due to the elusive nature of a definitive solution to scale ambiguity resolution, some researchers have explored sensor fusion to overcome some of the practices mentioned previously, such as bringing an object of known size into the environment or manually moving the camera through some sequence of poses. Nützi et al. (2011) details the authors efforts to solve scale ambiguity issues on a moving vehicle using the combination of a camera and an inertial measurement unit (IMU). Using simulated data, the method works remarkably well; using real world data from the IMU shows a high sensitivity to dynamic bias in the acceleration data from the IMU when calculating the scale of the environment. To illustrate the dynamic bias of an IMU, Figure 3 shows a plot of acceleration data from the z-axis of a stationary accelerometer. While this accelerometer, the MPU-6050, is a low-cost hobby-grade sensor and therefore may be more prone to bias, the data shown in the figure is indicative of bias even a commercial grade IMU may be subject to, albeit, to a lesser degree (VectorNav, n.d.). As the authors of the research describe in their work, this bias is difficult to quantify or express in a

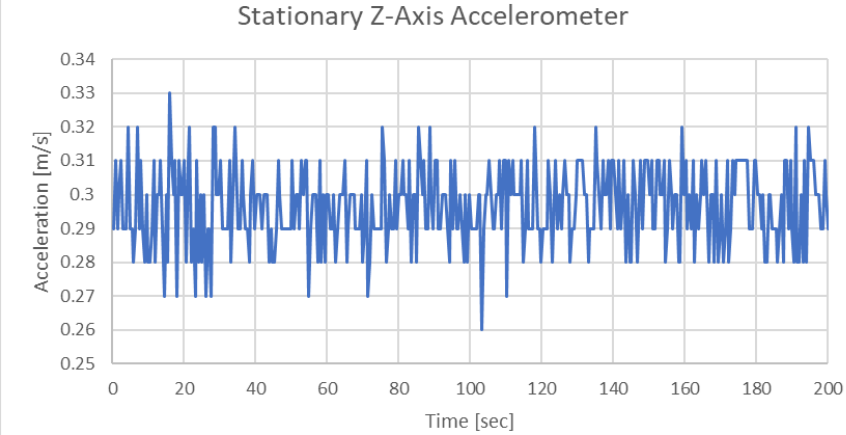


Figure 3: Stationary Data from Accelerometer Z-Axis

`

reliable model. Therefore, its impacts will almost certainly be observed in any scale estimation that is dependent on the fusion between an IMU and a vision sensor. Additionally, the method proposed by the authors relied on an Extended Kalman Filter (EKF) to filter the acceleration data from the IMU; this filter took at least 15 seconds to converge using real world data.

Investigation of Shutter Types

Consideration of the type of shutter working with imaging systems is a critical aspect of the design as they determine how the images are affected by motion and light. There are several different types of shutters to consider with the main division coming between mechanical and electronic shutters and different subdivisions of those characterizations.

Mechanical Shutters

As mentioned, there are several types of mechanical shutters such as the leaf shutter which is comprised of several overlapping blades built into the lens that open and close to allow light to reach the imaging sensor. The leaf shutter is easily recognizable as it is one of the more common illustrations of a camera lens. For example, searching for camera images in the Microsoft 360 suite yielded the icon in Figure 4. The advantage to using leaf shutters lies in its ability to use much faster shutter speeds in conjunction with a flash in comparison to other shutters. This is because the leaf shutter exposes the entire imaging sensor to light at the same time (Brown, n.d.). However, leaf shutters are mechanically complex and as such have a limited lifecycle before failure. Similar to leaf shutters, focal plane shutters are also mechanical shutters. Unlike leaf shutters, focal plane shutters are typically built into the imaging sensor rather than lens making them compatible with any lens that works with the imaging sensor. The focal plane shutter provides the functionality of a shutter by moving two coplanar curtains, separated by a small vertical slit, across the imaging sensor. The slit travelling across the sensor allows for light

Figure 4: Stock icon showing a leaf shutter

`

exposure to the sensor. As one might imagine, the motion of the slit across the imaging sensor can induce blurring, especially when imaging objects in motion (Brown, n.d.). Despite this, focal plane shutters are advantageous over leaf shutters when considering shutter speed, they are significantly faster than leaf shutters (Brown, n.d.). Figure 5 shows a diagram illustrating the functionality of a focal plane shutter. With both forms of mechanical shutter, the movement of the shutter can cause a phenomenon known as shutter shock which causes the imaging sensor to vibrate and cause the captured image to blur (Cromie, n.d.).

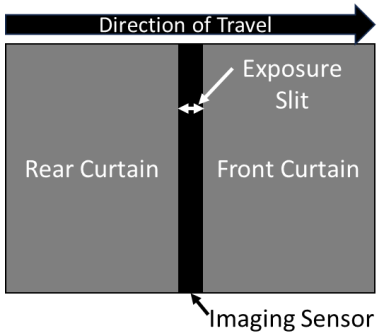


Figure 5: A diagram of how the focal plane shutter works

Electronic Shutters

An imaging sensor is usually comprised of a two-dimensional (2D) array of photoreceptors that capture incoming light with each element of the array corresponding to an individual pixel in the captured image. For certain imaging sensors, particularly sensors with small size formats, electronic shutters are beneficial in comparison to mechanical shutters. Electronic shutters function on a software level on the sensor. There are two primary types of electronic sensors, rolling and global shutters. Rolling shutters read a limited number of rows of the imaging sensor at a time until the entire sensor has been read. In similar fashion to the focal plane shutter, a rolling shutter can experience significant distortion and blurriness in environments with moving objects. Figure 6 illustrates how a rolling shutter works on an imaging sensor; the gray row is being read by the processor and once it has been read the next row will be read out. This process is repeated until all rows in the array have been read. In

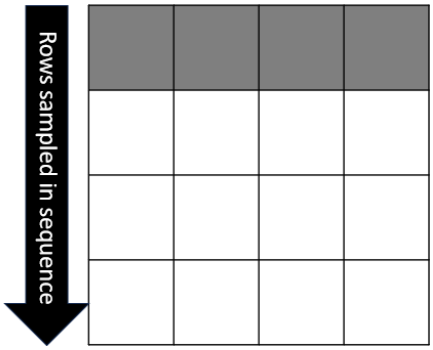


Figure 6: A diagram illustrating how a rolling shutter works

`

contrast to the rolling shutter, a global shutter samples all the elements of the array simultaneously. This sampling strategy is incredibly advantageous in dynamic environments as the sampling of the array occurs quickly such that dynamic objects have little opportunity to move while the array is

a) **b)**

Figure 7: a) A picture of helicopter's propellers taken with a global shutter camera. b) The same photo taken with a rolling shutter (Paul, 2016)**.**

being sampled. This results in little to no blurring or distortion in the captured image and is generally regarded as preferable to a rolling shutter. An example of the distortion induced by a rolling shutter can be seen in Figure 7b. With both types of electronic shutters, however, the inability to sync a flash with the shutter can become problematic. Flashes in photography are typically very bright, very quick strobes of light that are shorter in duration than that of an electronic shutter's sampling action (Nicholson & Summersby, n.d.). This requires sensors with electronic shutters to be properly illuminated from an external source.

Overview of Lenses

The simplest reduction of how a modern camera lens works is the thin lens theorem. This theorem can be described mathematically, as seen in Equation 1, where d_o describes the distance

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} \tag{1}$$

to the plane of focus from the lens, d_i describes the distance from the lens to the imaging sensor, and f describes the focal length of the lens. More practically, it can be illustrated using the image in Figure 8. Figure 8 shows a thin lens viewing a real-world object and projecting it onto an image sensor plane. The d_o, d_i, and f in the figure are the same as those detailed in Equation 1. It should be noted that, while they appear similar in Figure 8, d_o, and d_i are typically not similar distances.

`

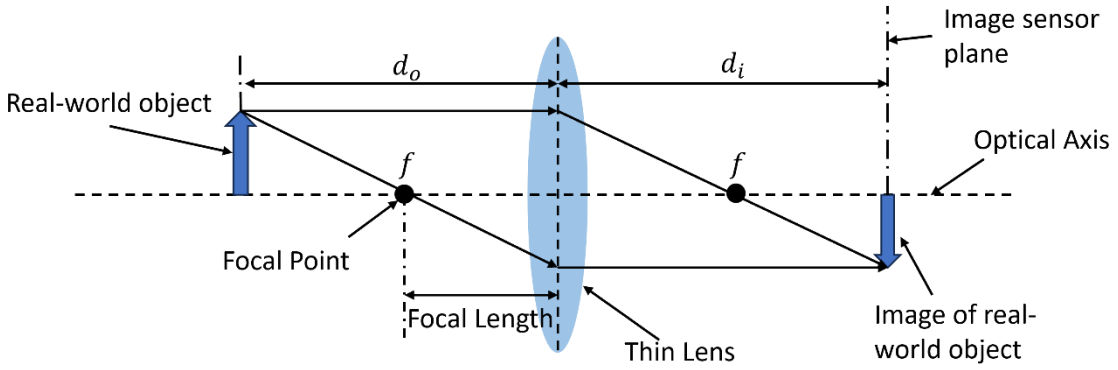


Figure 8: A visual representation of a thin lens (Photonics Media, n.d.)**.**

This representation of a lens is sufficient for most cases; however, modern camera lenses are often comprised of multiple thin lenses which adds some nuance to Equation 1. To better illustrate this, Figure 9 shows an illustration of a modern camera lens and its components as provided by a photography website. As is evident in Figure 9, there are often many thin lenses in a modern camera lens; an additional important feature of modern camera lenses in the aperture size of the lens as a whole. As can be seen in Figure 9, the lenses aperture is 400 mm. The aperture can also be represented as an "f-number" or "f-stop" which takes the form f/#, where the # represents some

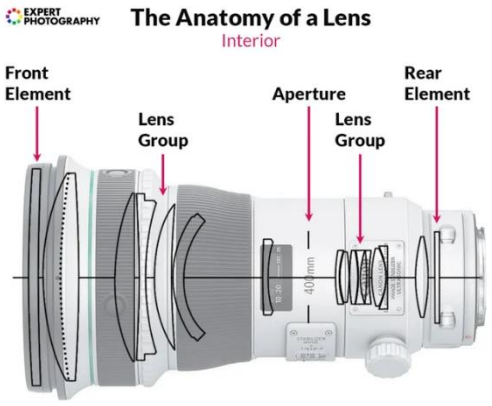


Figure 9: The anatomy of a modern camera lens *(ExpertPhotography, 2023)***.**

number. This can be calculated by dividing the focal length by the number represented by the # symbol (Nikon, n.d.). The aperture size plays a critical role in photography as it, similar to the shutter speed discussed previously, determines how much light is able to reach the imaging sensor,

`

as one would expect, a larger aperture allows a greater amount of light to reach the sensor when compared to a lens with a smaller aperture.

The Effect of Apertures on Depth of Field

Aperture size also directly impacts a phenomenon known in photography as the depth of field (DoF). The DoF can be described qualitatively by observing photographs; when one is taking a photo with a camera, the subject of the photograph is what one attempts to have as the most in-focus object in the photo. While focusing on the subject, whether it be manually or using some auto-focus algorithm, one may notice a small range of focal adjustments values in which the subject remains acceptably in-focus without distinguishable blurriness developing between adjustments of the focal length. To simplify, for each step adjustment of the focal length of a lens, there exists some range of distances from the lens in which a subject will be acceptably in-focus, this is referred to as the depth of field of a lens. When considering aperture, the DoF is reduced in size as the aperture size is increased. Inversely, the DoF enlarges with a smaller aperture size (Nikon, n.d.). This is because light rays extending from an object and passing through a lens, converge at a small circle rather than a dot. The size of this circle can be adjusted by adjusting the aperture size or the focal length of the lens (Mateer, n.d.). Graphics showing exactly what is meant by the circle of confusion is show in Figure 10 and Figure 11. In the figures, the object at the in-focus position emits light rays that enter the aperture, pass through the lens and then converge at

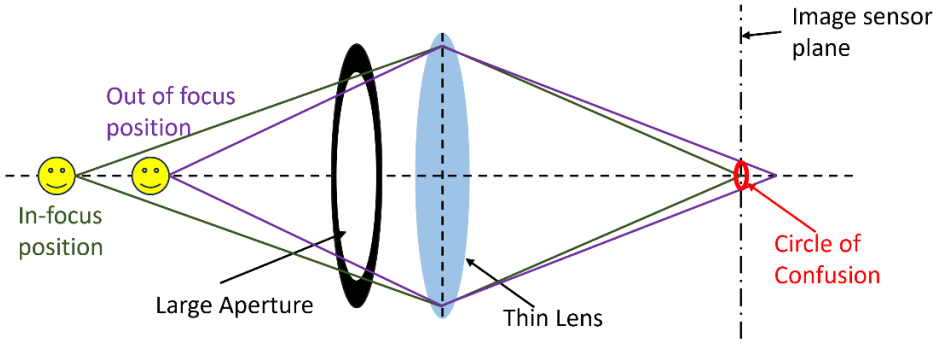


Figure 10: A circle of confusion caused by a large aperture.

`

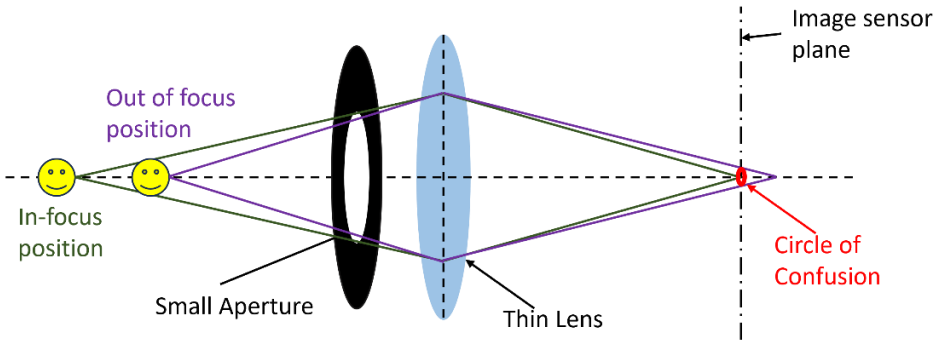


Figure 11: A circle of confusion caused by a small aperture *(Vision Doctor, n.d.)*.

the plane of the image sensor. The same object at an out of focus position in each figure emits similar light rays that pass through the lens and converge at a point behind the plane of the image sensor. The circle of confusion in the circle formed by these rays as they intercept the plane of the image sensor. A sufficiently small circle of confusion may be unnoticeable without serious effort and as such the object at an out of focus position may be clear enough in the image to be considered in-focus. As is evident in Figure 10 and Figure 11, the larger the aperture, the larger the circle of confusion; this indicates the distance between the in-focus and out of focus position in Figure 10 is smaller than the same distance in Figure 11, this distance defines the DoF. The circle of confusion becomes noticeable as it begins to grow larger than the size of the pixels on the image sensor array. As this happens, the same information begins to spill across a larger number of pixels and the image begins to become blurrier (Mateer, n.d.). In traditional photography, it is often desirable to have a larger DoF to capture as much information as possible in the photograph; when attempting to retrieve the distance to objects from an image, the inverse is true. The equation for the DoF can be seen in Equation 2 where A is the aperture size and C is the size of the circle of confusion.

$$DoF = \frac{2d_o AfC(d_o - f)}{A^2 f^2 - (d_o - f)^2 C^2} \quad (2)$$

`

Image Characteristics: Contrast and Intensity

An important discussion to be had when discussing image processing and image focus is the difference between intensity and contrast. Intensity refers to the amount of light a pixel represents (Nixon & Aguado, Chapter 2: Images, sampling, and frequency domain processing, 2012) whereas contrast relates to the disparity between high intensity and low intensity regions of an image, also referred to as the change in brightness (Nixon & Aguado, Chapter 1: Introduction, 2012). Both of these image characteristics can be used in a multitude of ways during image processing. In fact, measures of contrast, in conjunction with other metrics, have been used to determine image quality of images that have been altered in some way via an image processing filter very effectively, provided a reference image of the original image (Bhuiyan & Khan, 2015).

Scope of Paper

This work will focus on the creation of an initial prototype of a hardware-based solution to resolving depth from a monocular imaging sensor using depth from focus algorithms. The hardware components used in this prototype will be hobby-grade photography components. Exploration of monocular depth retrieval using deep learning methods falls outside the scope of this paper. This circumvents the issues associated with the initial setup of a deep learning algorithm as well as the additional computational power required for such algorithms.

Related Works

The concept of depth retrieval from a monocular imaging source has been previously studied by a diverse group of academics. In most of these works, laboratory grade or specialized photography equipment was used to retrieve the depth. Little work has been published on depth retrieval using hobby-grade components. Nevertheless, the same principles used on laboratory grade photography equipment can be applied to hobby grade components and, as a result, investigation of academic methods is a worthwhile endeavor.

`

Martel et al. (2018)

Martel et al. (2018), one of the primary works in this area of study, achieved real-time depth retrieval by manipulating the focus of a lens attached to a focal plane processor. A focal plane processor is a processor which performs analog operations and processing on the same chip that the image sensor array is on. This eliminates the need for an analog-to-digital converter which often bottlenecks image processing algorithms (Etienne-Cummings, Kalayjian, & Cai, 2001), (Fossum, 1989). This results in massive data throughputs and allows for very quick processing of images, thus aiding in the authors ability to perform real-time data acquisition from a camera. In addition to using a programmable focal plane processor, the authors of the paper also made use of a focus-tunable lens attached to the processor. This lens is a liquid-filled membrane with embedded current-driven voice coils which actuate the membrane based on the applied current (Optotune Switzerland AG, 2023). When the membrane actuates it changes shape to adjust the effective focal length of the lens without any mechanical inputs from an external user. As the control is entirely electronic, it is capable of actuating the lens at a high rate of speed, with the rise time of the input response of the lens equivalent to approximately 5 milliseconds (Optotune Switzerland AG, 2023). The complete lens setup used by authors is comprised of the focus-tunable lens farthest from the camera, a lens of negative focal length, and then an objective lens with a manually tunable lens. The negative focal length lens was required to bring the images from the focus tunable lens into the focal range of the objective lens. With the combination of a focal plane processor and a focus-tunable lens with a rapid response time, the authors were able to take several photographs in quick succession with photograph taken at its own unique optical power. This collection of photos is referred to as a focal stack. Once the focal stack was collected, each image was run through an algorithm which first blurred the image and then took the Laplacian of the image, in a process known as the Laplacian of Gaussian (LoG). The purpose of this was to first blur the image so that

`

noise in the image and artifacts such shadows or soft edges were blended with their surroundings. The resulting blurred image then only had the most prominent edges still visible in the image; the Laplacian of the image was then calculated to detect the edges. Once the LoG of each image was taken in the focal stack, the images in the stack were, in effect, overlayed on top of one another and composited together. The method for composition was to keep the pixel that maximized the response to the LoG and consider it the most in-focus pixel. By finding the most in-focus pixel at each possible pixel in the composite image, a depth cloud was constructed by correlating the in-focus pixel to its image's associated optical power and using Equation 1 to find the real-world distance to that pixel. There are some key issues associated with depth retrieval using this method and Martel et al. (2018) outline some methods for overcoming these challenges. One of the primary challenges with using a monocular imaging sensor with variable focal length to resolve depth of the environment is the range of real-world depths the sensor can sweep with good resolution. This is because the depth of field enlarges dramatically as the range of interest increases; this results in an increase of large error of depth values at larger ranges. Essentially, a point may be perceived as the most in-focus it can be at an image corresponding to an optical power associated with a distance from the camera of several meters and the DoF may be very large. For example, one of the lens configurations presented by Martel et al. (2018) had a DoF resolution of 0.225 m at a distance of 4.5 m from the camera. To combat this resolution challenge, the authors proposed optimizing the objective lens at the rear of the lens assembly specifically for the target range of the sensor. This is an effective solution because as the focal length of the rear objective lens increases so too does the maximum effective range of the lens apparatus (Martel, et al., 2018). However, doing so also reduces the field of view (FOV) of the overall image.

`

Martel et al. (2018) also mention in their work that they believe the proposed system will work on moving autonomous systems. This can be achieved by optimizing the speed of the focal stack collection by selecting the optimal number of images in the focal stack and prioritizing the speed at which they are collected, depth-retrieval can be done in real-time on a moving platform. Despite the promising nature of their work the authors further identify three major limitations of their proposed system. The first issue they discuss is what (Martel, et al., 2018) refers to as inpainting; this issue comes from the fact that the LoG method excels at finding edges in an image and nothing else. Therefore, very few pixels are detected as "in-focus" in each image. For example, when viewing a room with several objects in the FOV, Martel et al. (2018) observed that only 15% of the pixels in the final image were populated. This is because, essentially, only the outlines of objects were detected; they denote that determining the depth associated with missing pixels is referred to as inpainting in computer vision. The second issue discussed was that there was no denoising mechanism implemented in the algorithm. High intensity light sources may also produce a high LoG response which would detrimentally affect the depth estimation of the light source. Finally, due to the nature of the camera and lens system, it is impossible to acquire a continuous depth estimation for the environment as the depth is sampled at discreet intervals by way of taking a picture. (Martel, et al., 2018) proposes a potential solution, by way of assuming that depth values in between the depth retrieved from the focal stack is relatively close to the depth values around it, they refer to this solution as densification.

It is important to note after discussing the work presented in Martel et al. (2018) that the focal plane processing array used in the work is not a widely used instrument and, as such, there are few commercially available focal processor arrays (FPAs) (Teledyne FLIR, n.d.). This makes

`

reproduction of the work presented therein somewhat difficult due to the uniqueness of the hardware.

Nagahara et al. (2011)

While the work conducted by Martel et al. (2018) has been discussed at length, there have been other similar efforts from other authors. One such example is Nagahara et al. (2011) in which the authors make use of the same fundamentals of lens employed by Martel et al. (2018). Nagahara et al. (2011) also makes use of Equation 1; however, rather than varying the focal length, or optical power, of the lens, Nagahara et al. (2011) varied the value associated with the d_i term in Equation 1. This was accomplished by way of a linear micro-actuator attached to the lens which slid the lens away from the imaging plane, the setup used can be seen in Figure 12. This setup produces the same effect as the one described in Martel et al. (2018); but as Martel et al. (2018) points out, this setup suffers from some key setbacks. Namely, the difficulty in actuating the lens small distances,

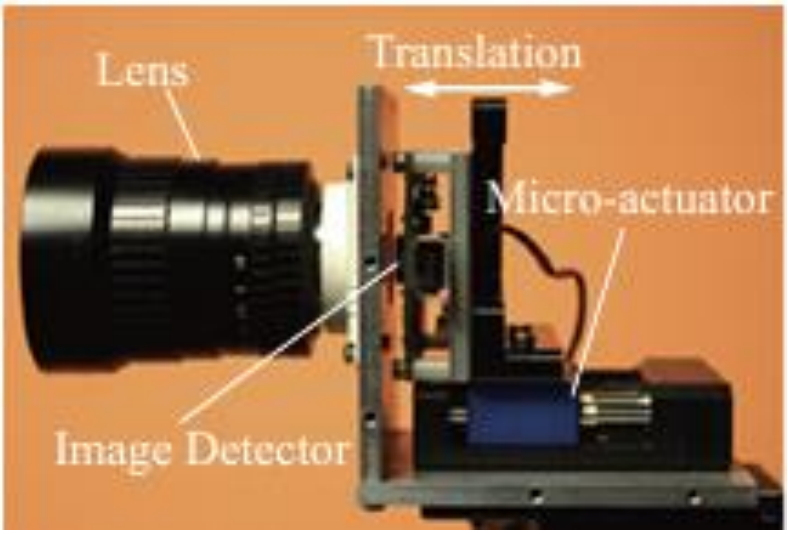


Figure 12: The lens setup used by Nagahara et al. (2011)

on the order of micrometers, quickly and accurately. This is where the benefit of using a current or voltage driven focus tunable lens, such as the one used in Martel et al. (2018), becomes beneficial. As is evident by Figure 12, the setup used by Nagahara et al. (2011) was custom

`

designed for their proposed solution which has an increased design complexity when compared to the design proposed by Martel et al. (2018).

Discussion on Consumer-Grade Photography Components vs. Lab-Grade Photography Components

In this work, extensive use is made of consumer-grade photography components; consumer-grade photography components differ from lab-grade photography components in that they are readily available to the average consumer. Empirically, one might distinguish consumer-grade photography components from lab-grade by the cost of the component. Typically, consumer-grade photography components are orders of magnitude less expensive than lab-grade and one does not usually have the opportunity to negotiate the price of consumer-grade photography components. Consumer-grade photography components are rarely custom made and, as a result, they are typically designed to interface with a large variety of other photography components by way of standardization.

Problem Statement

Given that perception on autonomous and semi-autonomous systems in a continuous challenge with no holistic solution, investigations into the real-world applications of vision systems such as the one proposed in Martel et al. (2018) are necessary. This work seeks to investigate the feasibility of depth retrieval from a small format consumer-grade camera system controlled by a small computer.

Methodology

The method to retrieve depth from a monocular camera system, presented in this work, evolved in many stages. After a brief preliminary evaluation of available options, it was decided the best course of action would be to use a focus-tunable lens similar to the one used in Martel et al. (2018). However, prior to proceeding directly to purchasing a focus-tunable lens, several forms of verification had to occur to ensure a proper baseline understanding of the fundamentals of lenses

`

utilized heavily in the reference literature. Initial investigations focused on developing algorithms to detect blurriness and eventually progressed to camera systems on sliding rails moving relative to the environment, to a geared lens, not unlike Nagahara et al. (2011), to finally integrating the focus-tunable lens into the system.

Overview of Fundamental Principles

Many electronically controlled focus-tunable lenses exist which can sweep through a large range of focus lengths, or optical power; these lenses are readily commercially available through retailers of optics equipment (Edmund Optics, n.d.). As a clarifying note for future discussions in this work, optical power is defined as the inverse of the focal length as described in Equation 3.

$$P_{otpical} = \frac{1}{f} \tag{3}$$

Equation 3 is a more concise method of describing the change of focal length. When observing the process outlined in Martel et al. (2018) in the context of Equation 1, one can see that the effect of changing the optical power while keeping the distance between the lens and the imaging sensor the same, i.e., not unscrewing the lens while changing the optical power, is to change the distance to the plane of focus. That is to say, each image in the focal stack corresponds to a unique distance from the camera at which the scene is acceptably in focus. To apply a Laplacian filter to an image is to find the second spatial derivative of the image (Fisher, S, Walker, & Wolfart, 2003) (Nixon & Aguado, Chapter 4: Low-level feature extraction (including edge detection), 2012). The filter is effective at finding edges as edges produce a high intensity value variation between pixels lying on the edge and adjacent pixels. Taking the first derivative of the image would yield a local maximum at pixels along an edge and as a result, the second derivative of the image is equal to zero at points of high intensity (OpenCV, n.d.). In practice, applying the Laplacian filter to the image is a simple task which typically involves applying a 3x3 kernel to the image. One of the most common kernels is the kernel seen below which is the same kernel used Martel et al. (2018).

`

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

This kernel is defined as the trace of a Hessian matrix. As mentioned in the discussion of Martel et al. (2018), the objective lens plays a key role in the DoF. This can be expressed mathematically using equations typically used to describe microscopes and telescopes. Equation 4 describes the magnification of a compound lens on a telescope with $f_{telescope}$ describing the focal length of the

$$Magnification = \frac{f_{telescope}}{f_{eyepiece}} \tag{4}$$

$$FOV = \frac{AFOV}{Magnification} = \frac{AFOV}{\frac{f_{telescope}}{f_{eyepiece}}} = \frac{AFOV * f_{eyepiece}}{f_{telescope}} \tag{5}$$

lens closest to the environment, typically the objective lens, and $f_{eyepiece}$ describing the focal length of the lens closest to the eyepiece used by a human observer (Hawkins, 2017), (Onah & Ogudo, 2014). The FOV of the image captured by the camera discussed Martel et al. (2018) can then be calculated by modifying Equation 4 to produce Equation 6. Magnification can also be defined using the d_o and d_i terms from Equation 1 to yield Equation 7. From Equations 4 and 7 one can see that the working distance, or the distance to the object in focus, is increased with a greater objective lens focal length. Thus, the compound lens system can be optimized for a selected range. However, as discussed when reviewing the work proposed in (Martel, et al., 2018), the objective lens typically is only optimized for a select range depending on the application of the system. Future works may consider a system with a controllable objective lens in addition to the primary lens.

$$FOV = \frac{AFOV}{\frac{f_{objective\ lens}}{f_{tunable\ lens}}} = \frac{AFOV\ x\ f_{tunable\ lens}}{f_{objective\ lens}} \tag{6}$$

$$M = \frac{d_i}{d_o} \tag{7}$$

`

Fundamental Methodology

The simplest reduction of underlying algorithm in all camera setups in this work is described Algorithm 1. Steps 2 through 6 are relatively invariant; however, Step 1 is determined

Algorithm 1: Fundamental Algorithm of this Work

1) Acquire a focal stack of images
2) Apply a LoG filter to each image in the stack
3) Composite the filtered images together by keeping the highest pixel value response to the LoG filter for each picture
4) Use the index of the image corresponding to the highest pixel value to determine the focal power of the image containing the pixel
5) Use the focal power to determine the working distance to the pixel
6) Optional: Create a depth map or point cloud from the working distance

by in the setup used to collect the focal stack. The method by which the images are collected in Step 1 is crucial and is the primary focus of this work.

Blur Detecting Algorithms

Prior to purchasing any imaging hardware, an effort was made to detect blurriness in an image to verify that such an algorithm would work on images captured in a focal stack. As is customary for evaluation of computer vision algorithms, this algorithm was tested on a standard test image, in this case the image was “Siamese_32.jpg” from the Oxford-IIIT Pet Dataset database (Parkhi, Vedaldi, Zisserman, & Jawahar, n.d.). The image depicts a Siamese cat sitting on a chair; it can be seen in Figure 13. While the database this image was retrieved from is typically used to train neural networks on object classification, the image works sufficiently well for the purpose of this work as it is feature-dense and has a discernible foreground and background. To begin work on a blur detection algorithm, the image was resized into a square, 512x512 pixel image and

`

converted to grayscale. From there, the image was discretized into a 4x4 array of even size cells, after this, 16 images were created, each with one unique cell that was left unblurred, with a random Gaussian blur kernel being applied to the cells in the image that were not left unblurred. The result was a batch of 16 images that looked similar to the image seen in Figure 14. Note: the cell numbers

Figure 13: Siamese_32.jpg from the The Oxford-IIIT Pet Dataset database *(Parkhi, Vedaldi, Zisserman, & Jawahar, n.d.)*

were not included in the produced test images, those are only for clarity purposes. As one can see in Figure 14, one cell up from the bottommost-right cell, was left unblurred in the creation of the image. Once these 16 images had been created, various methods of blur detection were performed on the images; for example, the root mean square (RMS) and Michelson contrast were calculated for each cell in each of the 16 images. Additionally, the LoG of each cell in each image was calculated to evaluate whether it could also be used as a metric of image clarity. This was done to

`

Figure 14: Siamese_32.jpg with cells of random blur applied to it

measure the effectiveness of each contrast measure and identify shortcomings of each method. Initial investigations used root mean square (RMS) contrast measurements to detect which cell in each of the 16 images was left unblurred. Following RMS contrast measuring, Michelson contrast was also used to attempt to detect the unblurred cells. Finally, the LoG of each cell in each of the images was taken to detect the unblurred cell. The equation for RMS contrast can be seen in Equation 8 where the image dimensions is a X-by-Y two-dimensional array, $f(x, y)$ is the normalized intensity at pixel (x, y) and μ is the mean of the normalized pixel intensities (Bhuiyan

$$C_{RMS} = \sqrt{\frac{1}{XY} * \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} [f(x,y) - \mu]^2} \quad (8)$$

& Khan, 2015). The equation for the Michelson contrast is expressed via the equation in Equation

`

9 where L_{max} and L_{min} refer to the maximum and minimum luminance in the image (NASA AMES Research Center, n.d.).

$$C_{Michelson} = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} \quad (9)$$

In this case, intensity is used as the measurement of luminance and Equation 9 can be rewritten in terms of intensities in Equation 10 with I_{max} and I_{min} being the maximum and minimum

$$C_{Michelson} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \quad (10)$$

intensities in the image. When calculating the contrast, each cell was treated as its own image; for initial investigations, MATLAB was used to evaluate the contrast measurements. When evaluating RMS contrast on Siamese_32.jpg as a method of evaluating image clarity, the algorithm was consistently able to achieve an average accuracy of 87.5% over 20 epochs when attempting to find unblurred cells. This translates to consistently missing two unblurred cells in a batch of 16 images. While this is acceptable for most cases, higher accuracy was desirable for use with depth resolution. To that end, the Michelson contrast metric was evaluated after evaluation of the RMS contrast metric.

The Michelson contrast was much less consistent than the RMS contrast with accuracy ranging from 75% to 87.5% per run and an average of 80% accuracy over 20 epochs. This is likely due to the fact that the Michelson contrast metric can be greatly impacted by a single spot of high or low intensity (Peli, 1990). Thus, as the image gets blurrier and the contrast between different regions begins to lessen, the Michelson contrast measurement becomes less and less effective.

Finally, the LoG of each cell was used to determine the clarity of the cell. This is the same method for finding image clarity used in Martel et al. (2018). It is important to note that the LoG does not measure contrast as the RMS and Michelson contrast equations do. Instead, the LoG is a

`

high-pass filter for edge detecting. This means that as edges become sharper in an image, i.e., the image is closer to being in-focus, the values produced by the LoG filter increase. Other edge detecting algorithms, such as Sobel and Canny, were not considered due to Sobel's sensitivity to noise and Canny's complexity (Sharifi, Fathy, & Mahmoudi, 2002). Of the three methods used for calculating the clarity of the cell, the LoG metric was by far the most accurate. Over 20 epochs, the LoG maintained an average of 100% accuracy. This is not to say the LoG metric is perfect in every situation and every image as it relies primarily on edges to detect clarity. Thus, situations such as staring at a blank wall or featureless room may reduce the accuracy of an algorithm using the LoG metric as a measure of clarity. A great example of one such situation also comes from the Oxford-IIIT Pet Dataset database, the image titled "Bombay_166.jpg" depicts a Bombay cat against a white background (Parkhi, Vedaldi, Zisserman, & Jawahar, n.d.). One of the generated test images using Bombay_166.jpg can be seen in Figure 15 and as one can see, cell 1 in the image has zero contrast in it which likely contributed to a lower average accuracy of 92.5% over 20 epochs using the LoG metric.

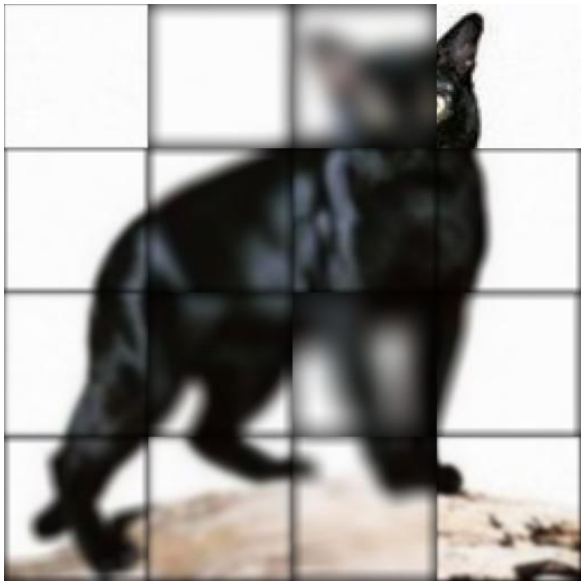


Figure 15: Bombay_166.jpg with cells of random blur applied to it (Parkhi, Vedaldi, Zisserman, & Jawahar, n.d.)

`

Interestingly, of the 20 epochs conducted on test images generated from Bombay_166.jpg, the LoG had the lowest overall average accuracy of the three metrics tested. The Michelson contrast performed well on the test images with an average accuracy of 93.75% over 20 epochs but the RMS contrast again outscored the Michelson contrast with an average accuracy of 96.88%. Table 1 has been made to succinctly capture the results of testing image clarity metrics on Siamese_32.jpg and Bombay_166.jpg. Despite its decreased performance on tests using Bombay_166.jpg, the LoG metric was selected as the method for finding image clarity and, eventually, determining focus of pixels in a focal stack. This was because of the robustness of the LoG operation. Not only is it capable of determining when an image is in-focus by finding when edges are maximally sharp, it also can quickly perform 2nd-order derivatives on the image to find the edges and localizing them correctly within the image (Bhairannawar, 2018), (Sharifi, Fathy, & Mahmoudi, 2002). Given the results in Table 1, enough confidence in the using the output of the LoG filter as a way to measure image clarity was established to proceed.

Table 1: A side-by-side comparison of image clarity metrics tested on the two images.

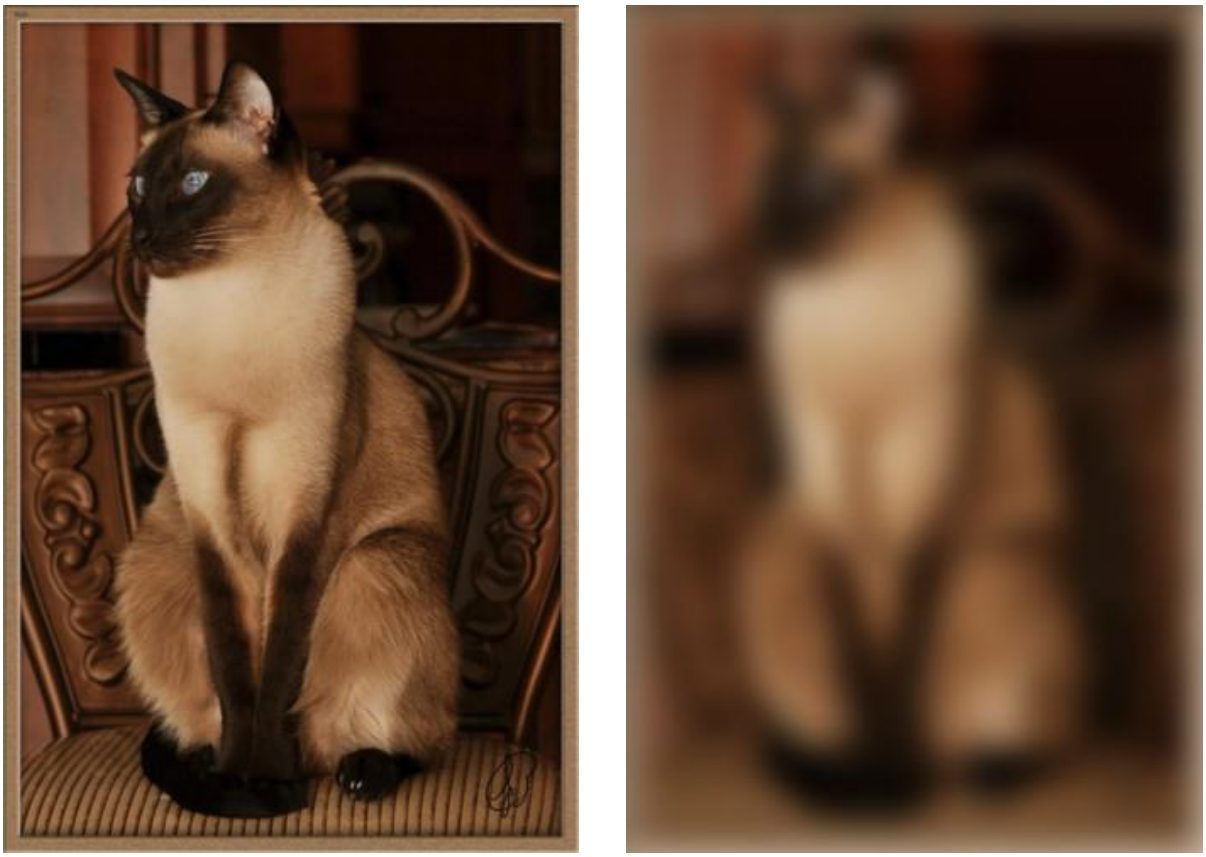
Image Name	Clarity Metric	Average Accuracy over 20 epochs with a batch size of 16
Siamese_32.jpg	RMS Contrast	87.5%
	Michelson Contrast	80%
	LoG	100%
Bombay_166.jpg	RMS Contrast	96.88%
	Michelson Contrast	93.75%
	LoG	92.50%

Simulating Changing Optical Power in Software

After evaluating the clarity metrics and selecting to proceed with the LoG metric, a modification to the cell-based test was conducted to confirm that the LoG would be able to detect in-focus planes in a focal stack. To verify this, the test images were subjected to a gradient of Gaussian blurs at a global level to simulate a focal stack taken while the focal power of the lens is being adjusted. This was accomplished by starting with a stack of the original test image and

`

applying a Gaussian blur with a small sigma, the standard deviation of the Gaussian blur, of 0.6 to the first image in the stack and linearly increasing the sigma of the blur for sequential images in the stack such that the first image in the stack was the least blurry and the last image in the stack was the most blurry. However, each time a simulated focal stack was created one image, selected at random in the stack, was left unblurred. It was this image that served to simulate an in-focus image that the LoG algorithm was seeking to find. Running one such test on Siamese_32.jpg and Bombay_166.jpg for 20 epochs yielded a 100% accuracy for both tests. An example of the first and last images in one of the focal stacks tested using the Siamese_32.jpg image as a base can be seen in Figure 16.



a) b)

Figure 16: The first and last image in a simulated focal stack

The sigma of the Gaussian blur was set to 0.6 times the index of the image in the focal stack; so, with a batch size of 16, the first image in the stack had a Gaussian blur with a sigma equivalent to

`

0.6 and the last image had a Gaussian blur with a sigma equivalent to 9.6. As one can see in Figure 16a, the image blur is nearly imperceptible whereas in Figure 16b, the blur is makes distinguishing the cat difficult.

This is most analogous to manually zooming a lens on a large format camera while trying to find the appropriate optical power. Once the focal stack had been successfully simulated in software and a basis had been established for the effectiveness of using the LoG to detect the most in-focus image in a focal stack, it became necessary to begin testing on real-world hardware.

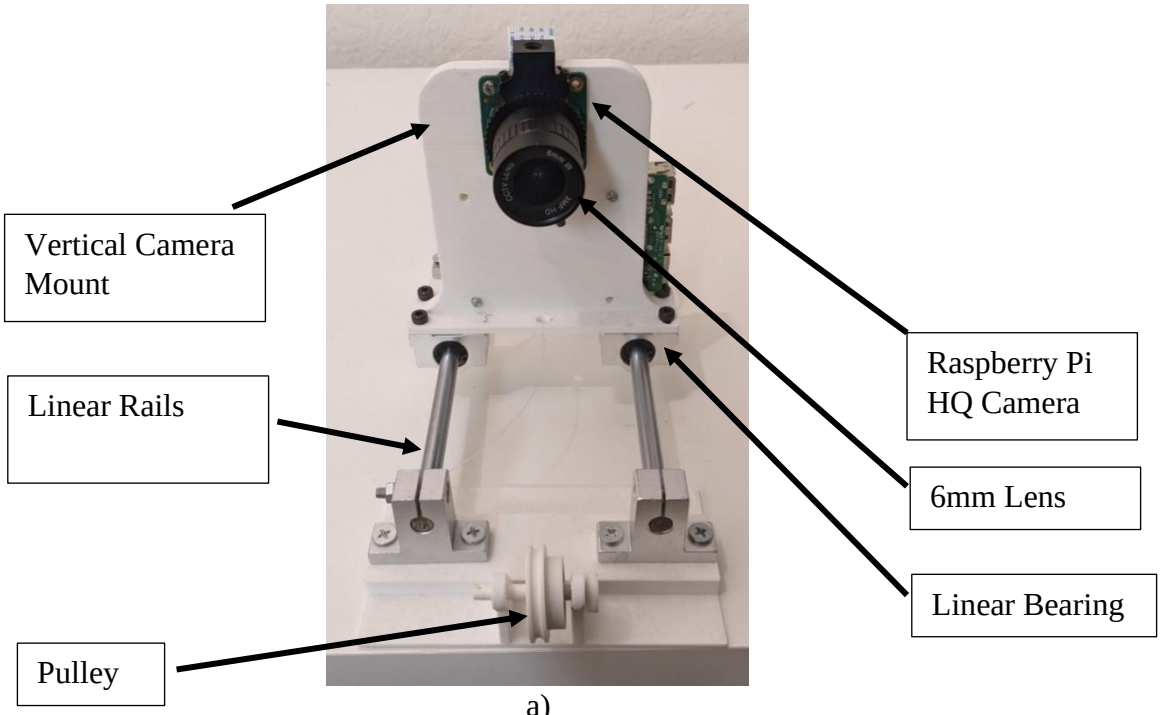
Linear Rail Hardware Setup

Due to its readily available nature and the ease of its use, the main computer used to drive the real-world cameras in this work was a Raspberry Pi 4 Model B. Two different compatible cameras were used in this work as well as two different lenses. Initial work began with the 12.3MP Raspberry Pi High Quality camera. However, this is a rolling shutter camera and as discussed previously, rolling shutters can produce significant blurring while capturing motion; additionally, it was thought to potentially be detrimental when varying the focal length and capturing images with a rolling shutter. Thus, a Raspberry Pi Global Shutter camera was ordered. Additionally, a 6mm 3MP lens was purchased for use with both cameras.

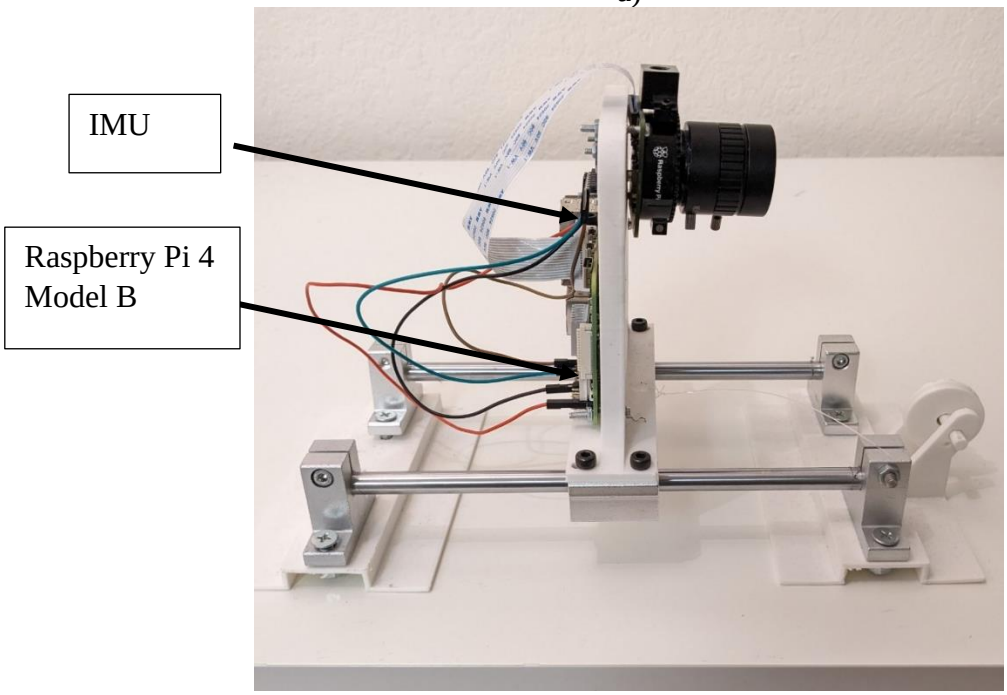
While awaiting the arrival of the global shutter camera, a simple test rig was developed to move the camera through the environment to begin investigations into the principles defined by Equation 1. The test rig consisted of a vertical camera mount fixed to linear bearings on a linear track, a weight was attached to the mount via a string draped over a pulley to induce motion in the rig and move the camera through the environment; additionally, an inertial measurement unit (IMU) was mounted to the vertical camera support to trigger the camera and measure its acceleration through the environment. A front and side view of the sliding rail test rig can be seen

`

in Figure 17, the weight drawn over the pulley is not pictured Despite not being pictured, the weight to pull the mount is of some importance as the camera setup can only be sampled so fast;



a)



b)

Figure 17: a) The front view of the sliding rail test rig. b) The side view of the sliding rail

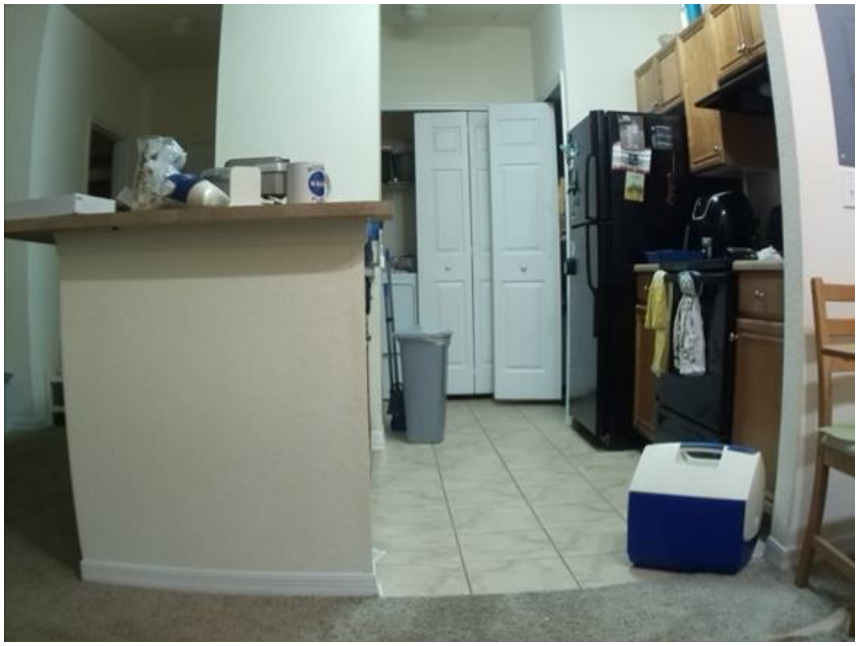
`

as a result, it is important to optimize the weight such that the camera moves at a rate at which sampling the desired number of images can be achieved before the end of travel. The test rig functioned by starting the camera as far from the pulley as possible and then releasing the weight attached to the vertical mount. As the weight fell, the IMU would detect a sudden spike in acceleration and begin to take pictures such that approximately 50 images were captured by the time the vertical mount reached the end of travel. When viewing this with respect to Equation 1, this is the equivalent of varying the d_o term in the sense that the camera is moving relative to the environment. As is evident from Figure 17, the sliding rail test rig is not a practical device as it would be difficult to automate and actuate rapidly enough to be useful. Additionally, when considering integration not a moving vehicle or platform, the sliding mechanism becomes a bigger challenge to design as it must move significantly faster than the vehicle it is integrated into. These challenges indicate a more robust design is required for real-world applications.

The sliding rail design does, however, serve useful to validate some of the principles discussed in Martel et al. (2018). Primarily the application of the LoG on a focal stack; by using open-source code provided by OpenCV, a Python script was used to perform Steps 1 through 3 in Algorithm 1. For most of the tests with the linear rail system, the camera was pointed towards a kitchen area, the original color image in the stack and the final image after the LoG had been applied to the focal stack can be seen in Figure 18. It is in Figure 18b that it becomes most apparent that a rolling shutter was used to collect the data; the streaking seen in some of the edges is indicative of this. However, the benefit of using the LoG is also apparent in Figure 18b as it is there that one can also observe the LoG's ability to resolve small details in images such as the pattern on the floor tiles as well as the edges of the carpet. This served as a proof of concept for actuating the lens of the camera and served as a steppingstone into the next phase of the

`

development. While moving the camera relative to the environment is a valid method for manipulating optical power according to Equation 1, moving the lens relative to the imaging plane is a far more compact method of achieving the same effect. Despite this, there are few commercial options for motorized lenses in the size range necessary for the Raspberry Pi cameras.



a)

b)

Figure 18: a) The first color image in the focal stack. b) The result of the LoG applied to the focal stack

`

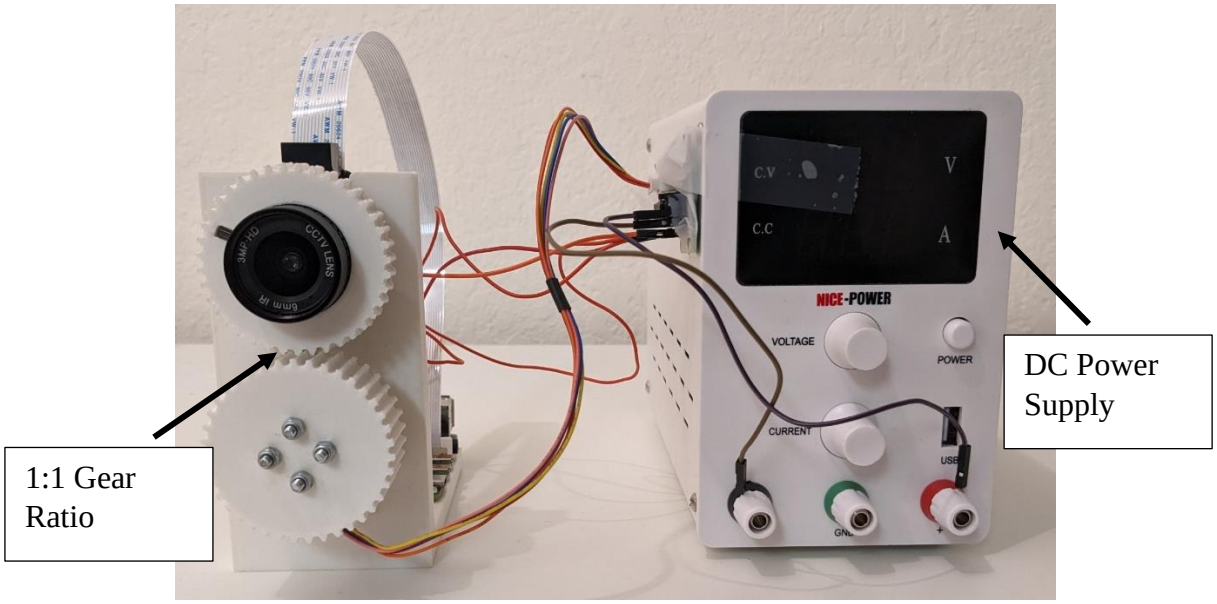
Geared Lens Setup

Taking inspiration from some hobbyist's efforts to create auto-focusing camera (Szczys, 2012), a gear system to actuate the lens of the Raspberry Pi Global Shutter camera was created so that the lens could be adjusted by the Raspberry Pi while it was capturing images. A system was created such that a 1:1 simple gear train driving the 6mm lens for the global shutter camera. The input gear is rigidly attached to a 28BYJ-48 Stepper Motor which is controlled by the Raspberry Pi 4 via inputs to the ULN2003 Stepper Motor Driver. The driver also requires 5-12V of DC power supplied to it to drive the stepper motor. The stepper motor has a stride angle of approximately 0.088° (Kiatronics) and the lens, as it is a C-Mount lens, has a pitch of 32 threads per inch which is approximately equivalent to a 0.794 mm pitch. Images of this system can be seen in Figure 19. The geared lens setup functioned by varying the d_i term in Equation 1; the 6mm lens for the Raspberry Pi camera is a fixed focal length camera, meaning the f term in Equation 1 is invariant; in fact, the 6mm in the name of the lens refers to the focal length of the lens. Given that the camera is stationary, unlike the linear rail test rig, the d_o term is not a term that is manipulatable by the camera system. So, in order to adjust which plane in the environment is in focus, the rear of the lens must be moved relative to the imaging plane. Given this, the step angle of the stepper motor, and the pitch of the threads for the C-mount lens, an expression can be created to describe the change in d_i based on the number of steps input from the stepper motor. This expression can be

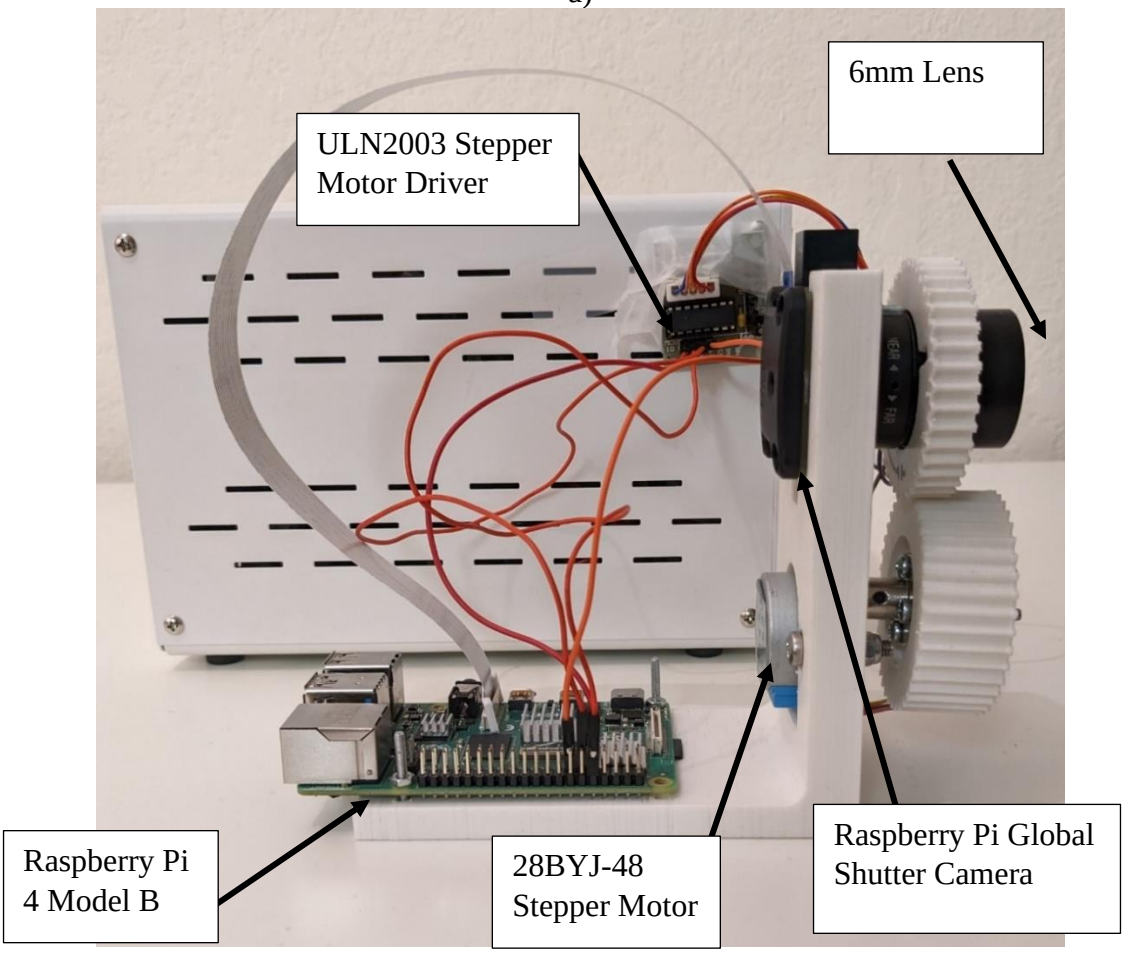
$$\Delta d_i = \frac{n_{steps}}{4096} * 0.794mm \tag{11}$$

seen in Equation 11 where Δd_i refers to the change in d_i, n_{steps} is the number of steps input from the stepper motor, 4096 is the total number of steps required for the stepper motor to make one full revolution, and $0.794mm$ is the pitch of the camera lens. From Equation 11, it is clear that a stepper motor capable of exceptionally small step sizes would be ideal for this scenario.

`



a)



b)

Figure 19: a) The front view of the geared lens setup. b) The side view of the geared lens setup

`

Algorithm 2: Auto-focus of the geared lens setup

Input: $im_{init}, n_{steps} = 0, LoG_{init} = LoG(im_{init})$	(Capture an initial image, the stepper motor has moved 0 steps, get the LoG of the initial image)
$n_{steps} = n_{steps} - 1000$	(Move the stepper motor 1000 steps in some direction)
$LoG_{curr} = LoG(im_{curr})$	(Get the LoG of the current image)
If: $LoG_{curr} > LoG_{init}$	(If the LoG of the current image is greater than the initial log)
$LoG2_{curr} = 0, LoG1_{curr} = LoG(im_{curr})$	(Initialize two LoG variables, one at zero and one at the LoG of the current image)
While: $LoG1_{curr} > LoG2_{curr}$:	(While the LoG1 is increasing)
$LoG2_{curr} = LoG1_{curr}$	(Set the LoG2 variable equal to LoG1)
$n_{steps} = n_{steps} - 20$	(Drive the stepper motor by 20 steps)
$LoG1_{curr} = LoG(im_{curr})$	(Set LoG1 equal to the LoG of the current image)
end	(The image is focused)
Else:	(If the image got blurrier after moving the motor 1000 steps)
$n_{steps} = n_{steps} + 2000$	(Move the motor the opposite direction by 2000 steps)
$LoG2_{curr} = 0, LoG1_{curr} = LoG(im_{curr})$	(Initialize two LoG variables, one at zero and one at the LoG of the current image)
While: $LoG1_{curr} > LoG2_{curr}$:	(While the LoG1 is increasing)
$LoG2_{curr} = LoG1_{curr}$	(Set the LoG2 variable equal to LoG1)
$n_{steps} = n_{steps} + 20$	(Drive the stepper motor by 20 steps)
$LoG1_{curr} = LoG(im_{curr})$	(Set LoG1 equal to the LoG of the current image)
end	(The image is focused)
end	(End algorithm)

Using this setup, an auto-focus algorithm was created which would serve as the basis for the remainder of this work. Similar to simulations prior to any hardware setup, the LoG was used to determine image clarity in the geared lens setup. The algorithm for auto-focusing the lens can be seen in Algorithm 2. Essentially, the algorithm guessed a direction to spin the lens by 1000

`

steps of the stepper motor and monitored if the image became blurrier, if not, it kept spinning the lens in that direction until the image was maximally clear, checking every 20 steps of the stepper motor. If the image did get blurrier, the algorithm spun the lens 2000 steps in the opposite direction and then kept spinning the lens in the same direction until the image was maximally clear, checking every 20 steps of the stepper motor. It is interesting to note, with small changes between checks, occasionally the algorithm will incorrectly pick the direction to spin the lens, especially when the initial lens is very blurry. It is as this point, a human must manually seed the initial lens position such that the image is nearly in focus. This is due to the lack of absolute positioning mechanisms in the setup, i.e., if one were to power cycle the system the driver nor the Raspberry Pi would know the absolute position of the lens relative to the imaging plane, nor would they know if it had been moved while the system was powered off. This is a critical shortcoming of this geared lens setup as it requires any testing of the system to be initialized by focusing on a plane of known distance away from the camera. The most obvious solution is to point the camera at a reference on the sliding rail gantry; however, this would impede the perception of the camera. So, to remedy this, a 152.4mm x 152.4 mm square was printed on a piece of paper and taped to a wall near the table the setup was resting on. This added a feature to the wall the LoG algorithm could detect and made auto-focusing to a known depth rather simple as once the camera had been focused on the square, the distance between the wall and the front of the camera could be measured, yielding the d_o term from Equation 1. Given that the focal length was fixed at 6mm, the lens could then be pointed in any direction and the most in-focus plane was known. From there the lens could be moved via the stepper motor while pictures were taken at discrete, known intervals at which the Δd_i could be calculated and summed to the previous intervals' Δd_is such that the total change in d_i could be calculated and, as a result, the d_o at which the in-focus points lay could also be calculated. Before

`

proceeding, it is critical to discuss another limitation of the geared lens setup. If one were to solve Equation 1 for d_o in terms of d_i, it would yield the expression seen in Equation 12, which is an

$$d_o = \frac{1}{\frac{1}{f} - \frac{1}{d_i}} \tag{12}$$

asymptotic function. This presents a unique challenge for manipulating the d_i term; to help

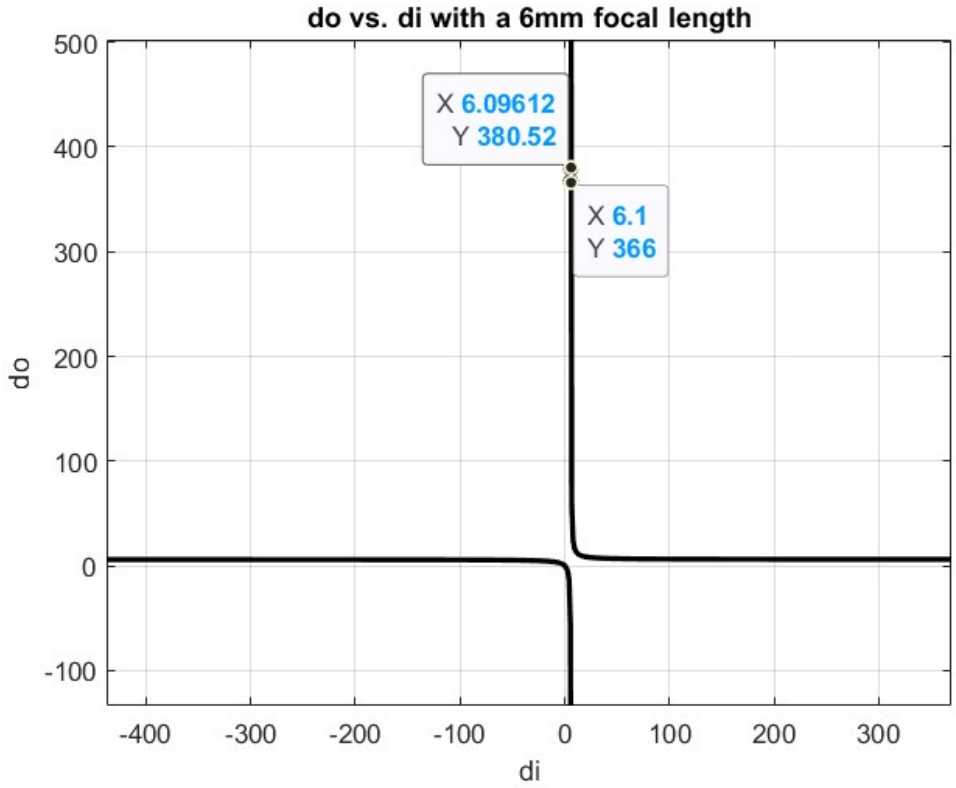


Figure 20: The graph of Equation 12

illustrate the challenge, the graph of Equation 12 with f equal to 6mm is shown in Figure 20. The two points on the graph show the difference, in millimeters, between the $d_o$value at a d_i equivalent

$$d_o = \frac{1}{\frac{1}{f} - \frac{1}{d_i}}$$

to 6.1mm and a d_i where the lens has been moved 20 steps of the stepper motor closer to the imaging plane, which is equivalent to 6.096mm. The difference between the d_o values is approximately 14 mm. The issue with the asymptotic nature of the function lies in the fact that this difference does not linearly scale based on the number of steps taken by the stepper motor.

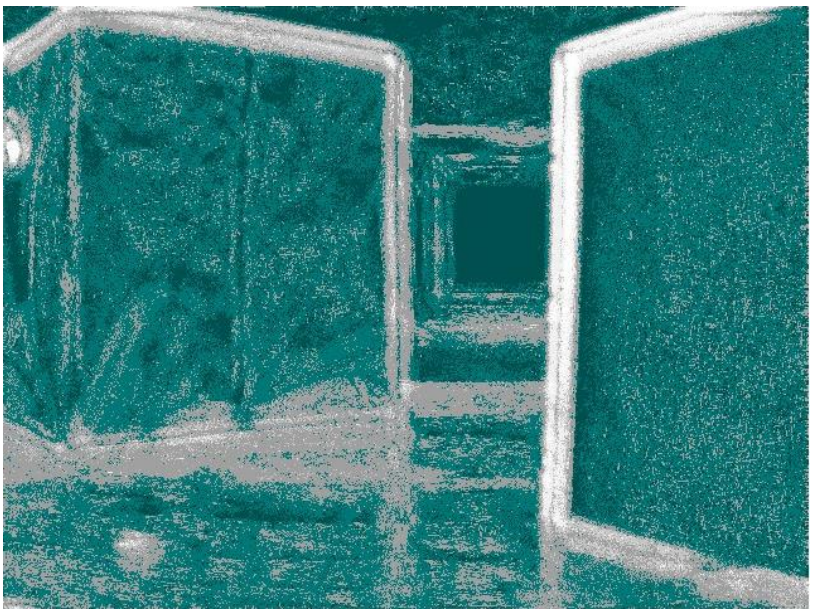
`

Meaning, if the initial d_i measurement in the graph in Figure 20 were closer to the focal length, 6mm, then a 20 step movement of the lens closer to the imaging plane would yield a greater difference in the d_o values. Due to backlash in the gears in the geared lens setup, it was noted empirically that it was difficult to move the camera lens at smaller increments than 20 steps. Additionally, the phenomenon denoted above would quickly reach a point where the change in d_i to yield a reasonably small difference in d_o values would be so small that it would not be achievable within a single step of the stepper motor. Issues such as this could be mitigated, to a degree, via half-stepping or micro-stepping; however, such explorations fall outside the scope of this work. The fact that small changes in d_i, when d_i is near the focal length, yield large changes in corresponding d_o values also effectively limit the maximum distance from the camera that the depth can reasonably be inferred. This distance is sometimes called the working distance. For small focal length lenses such as the ones used in this work, this distance is relatively short at less than 5 meters.

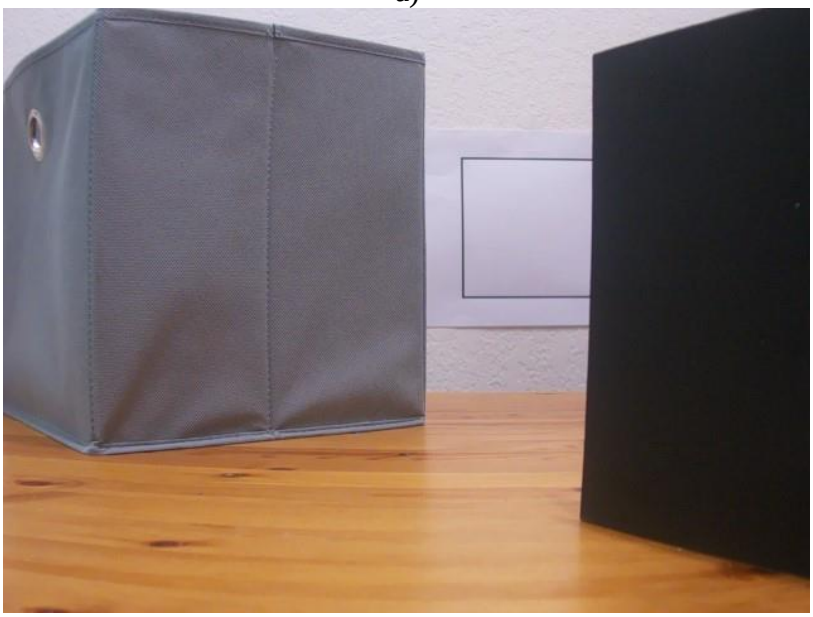
With a good understanding of the geared lens setup, testing began with it at short ranges. Tests were initially conducted in a densely cluttered room with a length of nearly 8 meters. It was during these tests that the limitations of the lens began to materialize as error in the data and, for the sake of brevity, they have been omitted from this work. As mentioned above, the geared camera lens setup required some initialization to properly begin its depth estimation. This process is described above; once the camera had been focused and the distance to 152.4mm x 152.4mm square the camera had been focused to had been measured. The initial d_i was calculated using Equation 1; the camera was then positioned in a desired location such that it was pointing at a cluttered environment. From there, the camera's lens was spun to focus on progressively farther planes and, at each plane, an image was captured. The LoG was taken of each image was taken

`

and the highest responding pixels to the LoG algorithm were placed in their respective positions in the final composite image. One such composite image can be seen in Figure 21a in which the values for all pixels in the grayscale image have been multiplied by 20 to increase the clarity of the image. As one can see in the image, there is a significant amount of noise in the image; one



a)



b)

Figure 21: a) The composite image of the LoG of each image in the focal stack, multiplied by a factor of 20. b) A photo of the environment.

`

may observe that the edges of the objects in the environment, seen in Figure 21b, tend to compound and thicken in the composite image. This may be an artifact of a non-constant magnification as the lens was moved via the gear or it is possible that it was induced by some small wobble in the lens as the lens was manipulated by the gear. When filtering for the noise and mapping the pixels to their corresponding depth via a color map, the image seen in Figure 22 is obtained. The areas of dark blue are intended to represent areas nearest to the camera whereas green is supposed to be indicative of areas farther away. As one can see, the depth mapping is largely error prone and susceptible to noise. This is likely caused by large overlaps in the depth of field of the images in the focal stack which would cause the same sections of an edge to be in-focus in multiple images and thus be incorrectly binned into the wrong depth values multiple times.

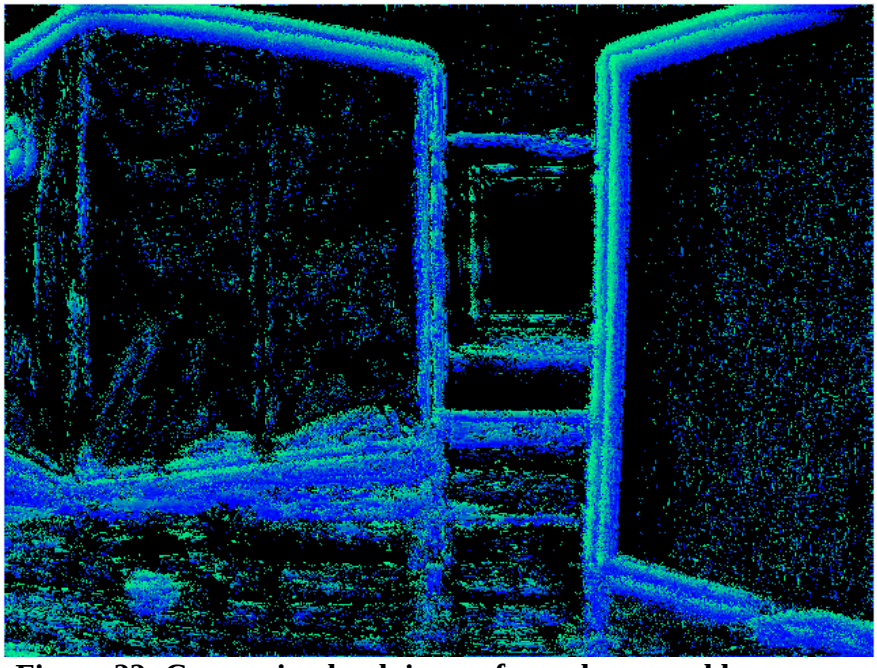


Figure 22: Composite depth image from the geared lens setup.

While there were significant errors associated with using the setup, the geared lens setup did however yield valuable insight into the process for use with a liquid focus-tunable lens.

`

Liquid Focus-Tunable Lens

Using the same vertical mount as the geared lens setup, a 12mm, f/6, Liquid Lens Cx Series Fixed Focal Length Lens from Edmund Optics was affixed to the Raspberry Pi Global Shutter camera. The liquid lens in this setup was a Corning Varioptic A-25H0 Variable Focus Lens. A diagram of the lens, provided in the manual by Edmund Optics, can be seen in Figure 23. The figure depicts a fixed focal length lens without the liquid lens; the liquid lens slots into the opening labelled "*Easy Access to Integrate Accessories Like Liquid Lenses, Filters, and Aperture Stops".*

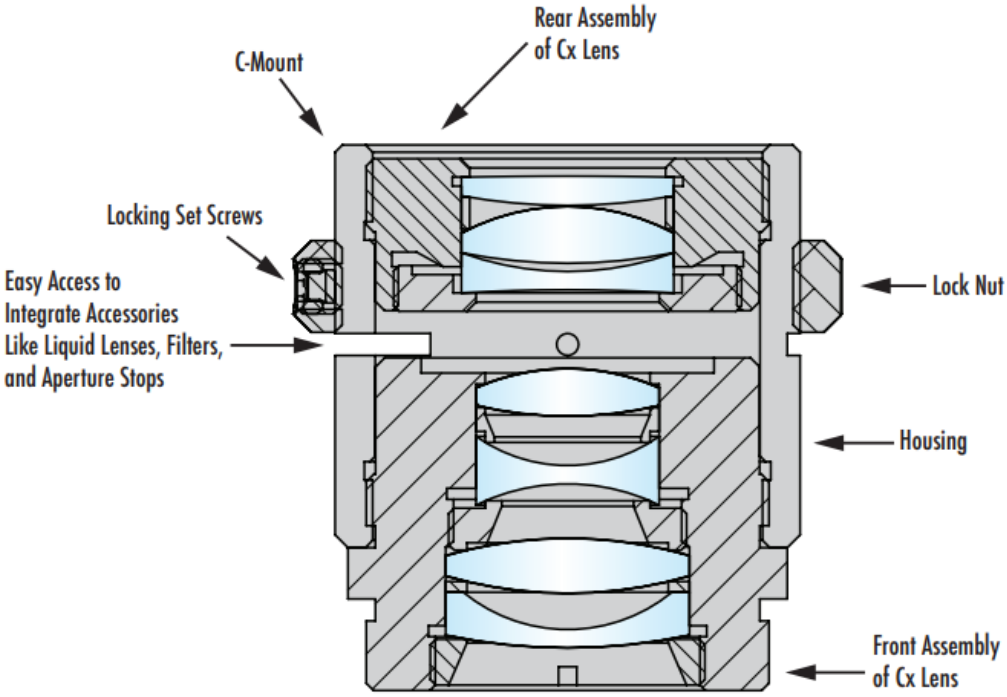
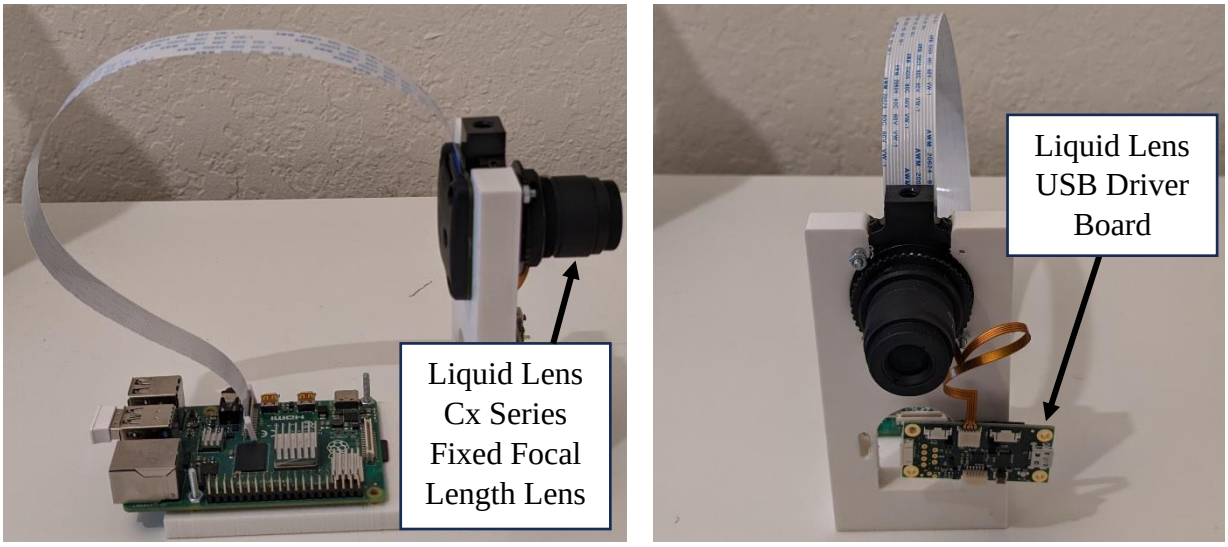


Figure 23: A diagram of the Cx Series Fixed Focal Length Lens provided by Edmund Optics (Edmund Optics)**.**

The base lens is a 12mm focal length lens that can be mounted into the global shutter camera. The liquid lens is comprised of a liquid-filled meniscus that can be actuated via a voltage controller driven by a software package provided in the development kit. Changing the voltage applied to the lens changes the optical power of the lens, much like the current driven lens used in Martel et al.

`

(2018). The complete camera setup can be seen in Figure 24. As the figure depicts, this setup is the most compact of all setups thus far. Future design can increase the compactness of the design by better packaging the driver board of the lens. For efforts detailed in this work, the liquid lens' optical power was adjusted manually using a Windows application that came with the lens. However, the lens also came with a driver board that could be used to drive the lens via code automatically from a microcontroller. In future works, this could allow for an auto-focus algorithm, similar to the auto-focus algorithm created for the geared lens setup, to be created for the liquid lens setup.



a) b)

Figure 24: a) The side view of the liquid lens setup. B) The front view of the liquid lens setup

This camera system setup was placed in an environment containing checkerboards on the ground and the 154.2mm x 154.2mm reference square in the background. An all-in-focus image of the environment can be seen in Figure 25 where the checkerboard progresses away the camera and at the wall on which the reference square is hung. Using this environment, a focal stack was collected by varying the optical power of the lens manually and collecting images at focal powers corresponding to every 0.01V increment in a voltage range that was optimized for the depth of the

`

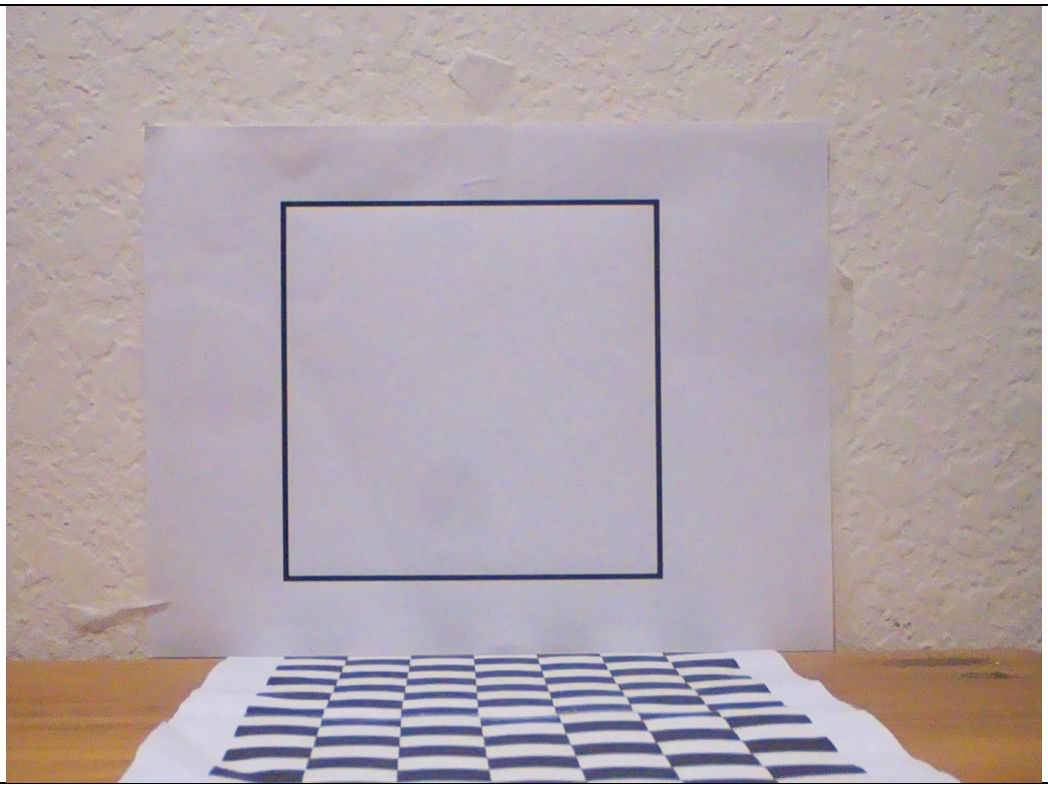


Figure 25: An all-in-focus image of the checkerboard environment

environment. From experimentation with the lens kit, an equation relating the working distance and the voltage applied to the lens was created. This equation is the slope of the line seen in Figure 26 and can be used to calculate the working distance at a voltage applied to the liquid lens. The

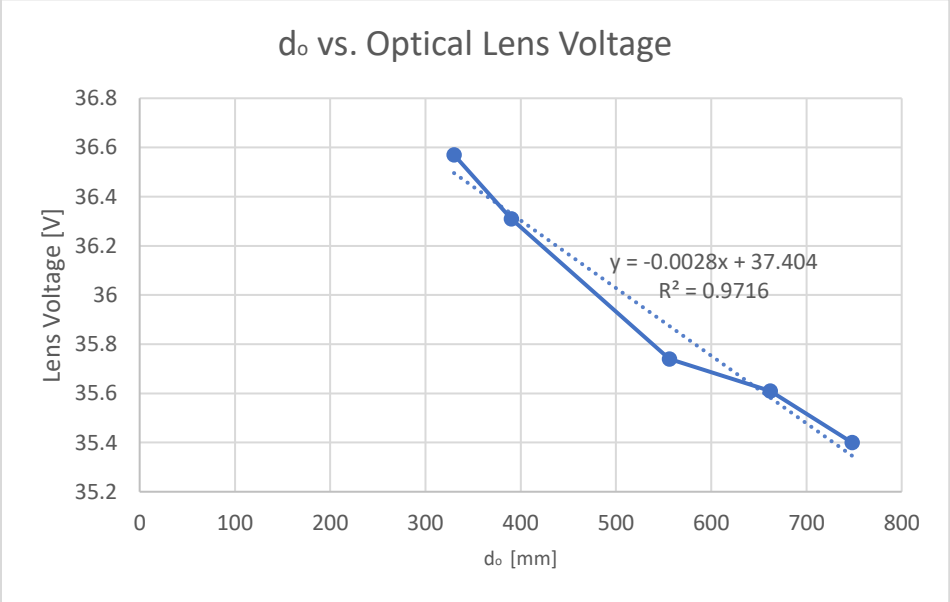


Figure 26: d_o vs. Optical Lens Voltage

`

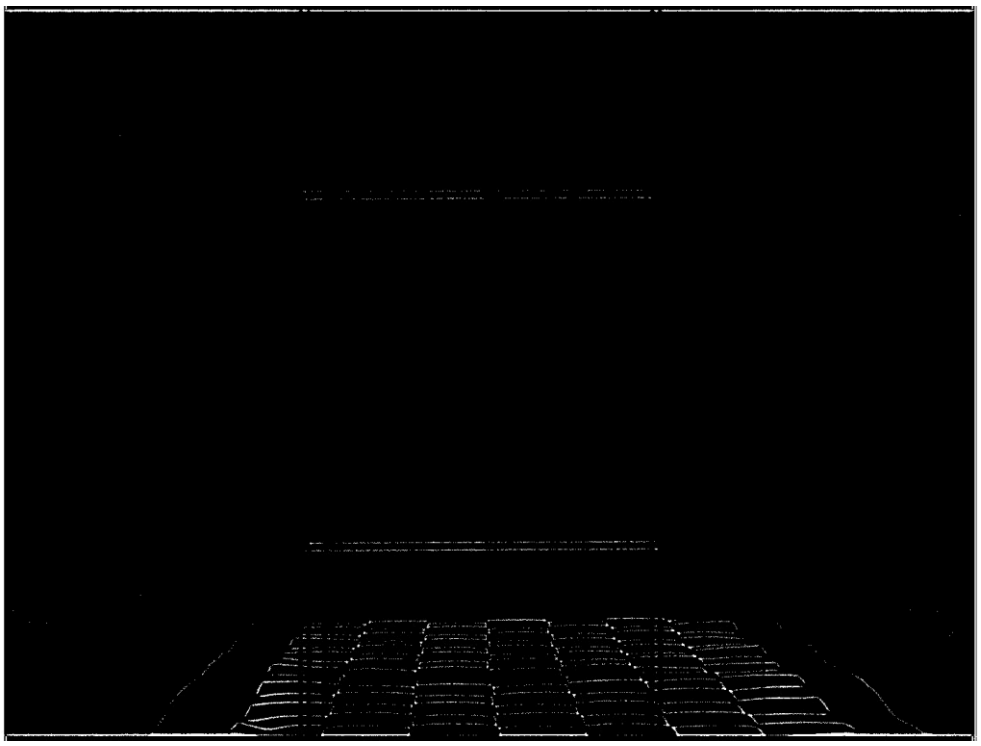
images collected in the focal stack then had the LoG algorithm applied to them to isolate the in-focus pixels in each image, these images were then composited together to get the result of all the images' response to the LoG. Once this was done, the index of the image associated with each pixel's highest response to the LoG algorithm was used to assign a color value to pixels that correlated to their depth as it relates to the index of its original image.

Results

As stated previously, the environment used to evaluate the depth from focus can be seen in Figure 25. In that environment, the 152.4mm x 152.4mm reference square was 1065mm away from the front face of the lens housing. In the first image, the far distances, seen in a, of the environment are most in focus. This is evident from the top border of the reference square in the background being detected by the LoG algorithm. Notably, there is a significant amount of noise stemming from the near distance ranges in this image. In b, the mid ranges in the environment are most in focus; this image comes from the middle of the focal stack. One can observe the nearly noiseless image in the areas of the image corresponding to greater depth ranges. However, this is not true of ranges closer to the camera as b still has significant noise from the near ranges. Finally, the nearest ranges of the environment are in focus in c. Much like b, the areas of the image in c corresponding to ranges greater than those that are in focus are largely noiseless. A composite depth image from a focal stack can be seen in Figure 28; this depth image is also prone to error but less so than the depth image seen in Figure 22. To produce the depth map in Figure 28, a simple thresholding technique was applied to remove the noise produced by weaker responses to the LoG algorithm. This proved effective at removing significant amounts of noise in regions where little to no edge features were present. The depth cloud seen in Figure 28 was projected onto an overhead 2D image of the environment and can be seen in Figure 29; as one can see in the figure, the noise

`

seen in Figure 28 greatly impacts the accuracy of the point cloud projection. However, it is important to note that the point cloud does have many linear discontinuities in it that appear to be



a)

b)

`

c)

Figure 27: a) The far distances in the environment are in focus, b) The mid-range distances are in focus. C) The close-range distances are in focus.

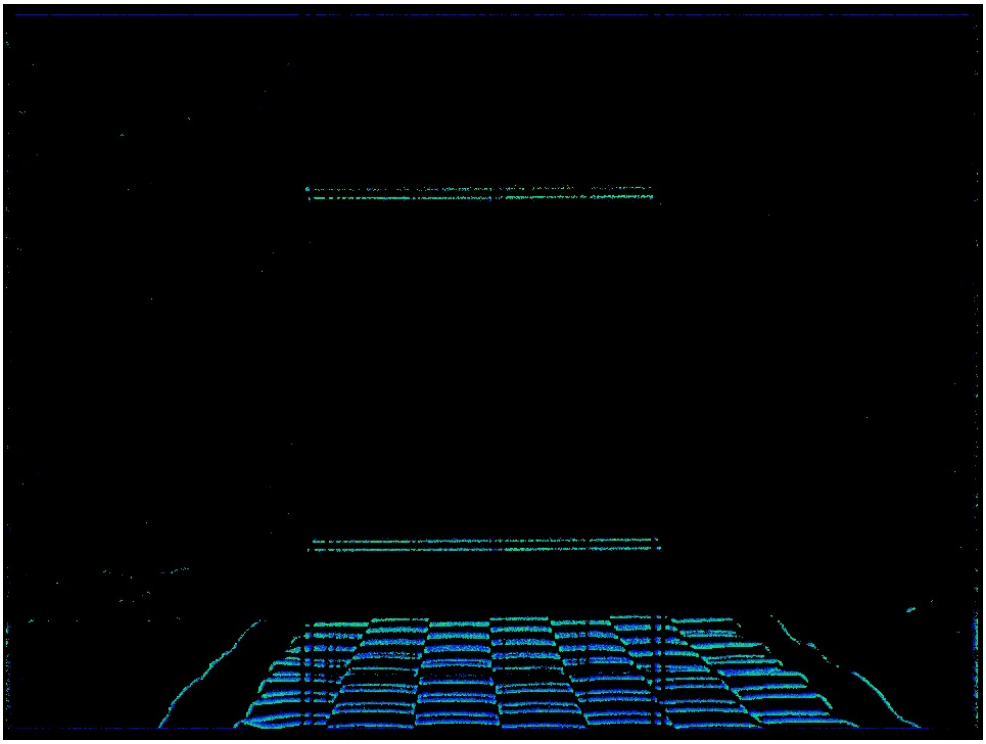
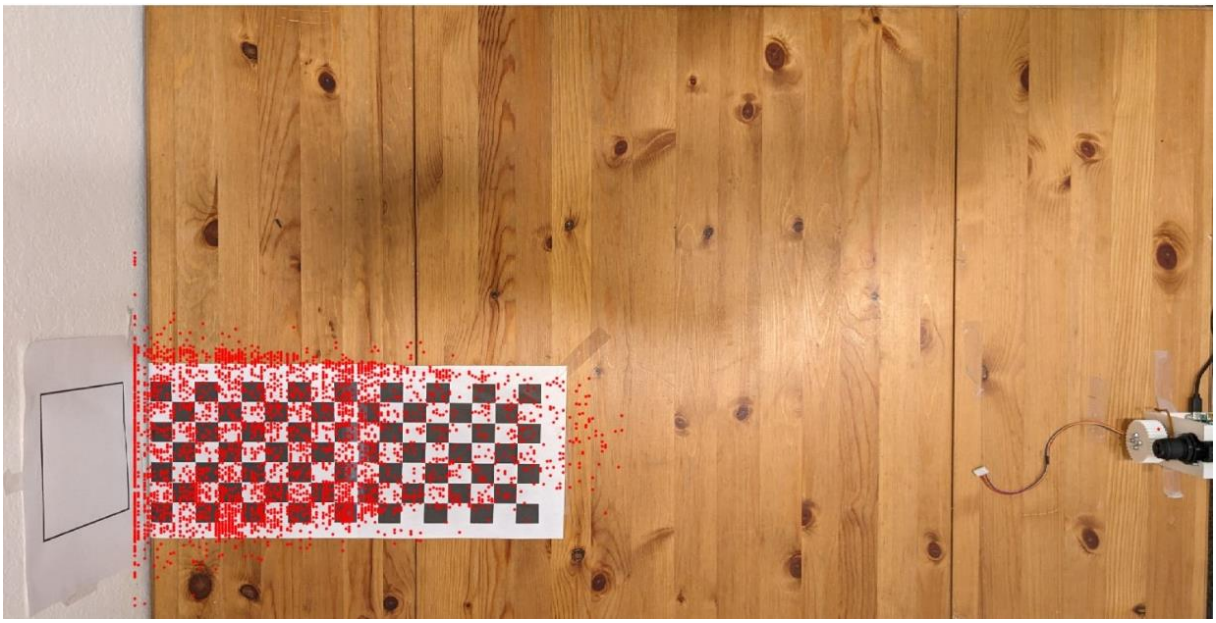


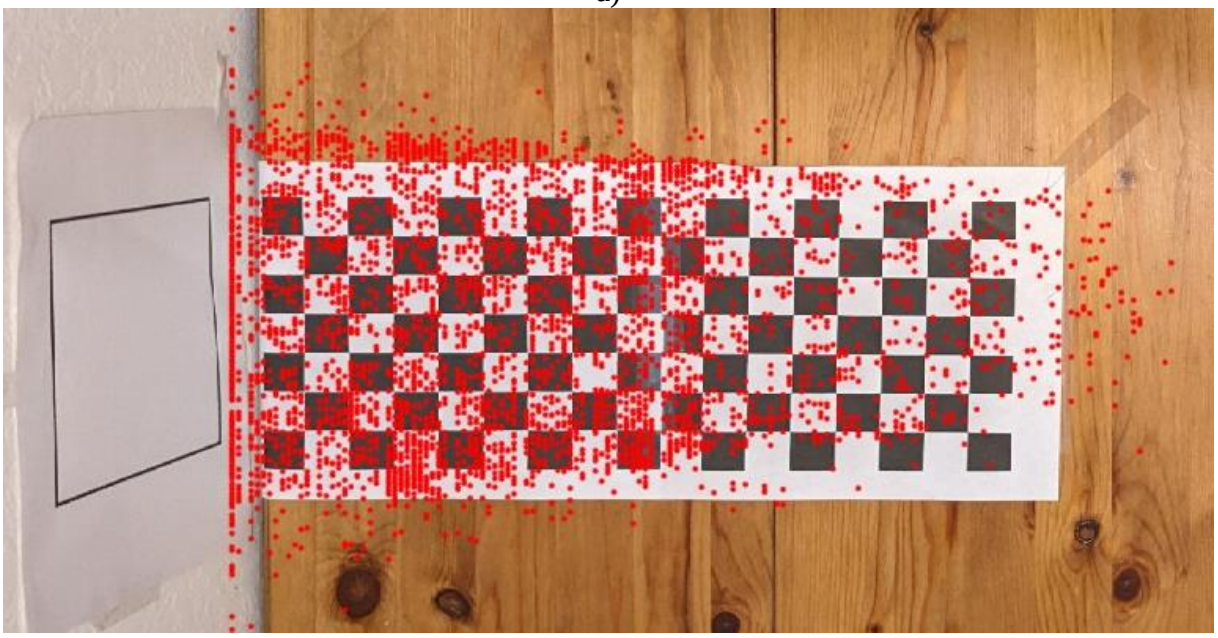
Figure 28: Composite depth image from liquid lens setup

`

close to the edges formed by the squares in the checkerboard. These discontinuities can be better seen in the image in Figure 30. The fact that the depth map and point cloud contain so much noise indicates a need for improvements to the system.



a)



b)

Figure 29: A projection of the depth map seen in Figure 28 onto an overhead view of the environment.

`

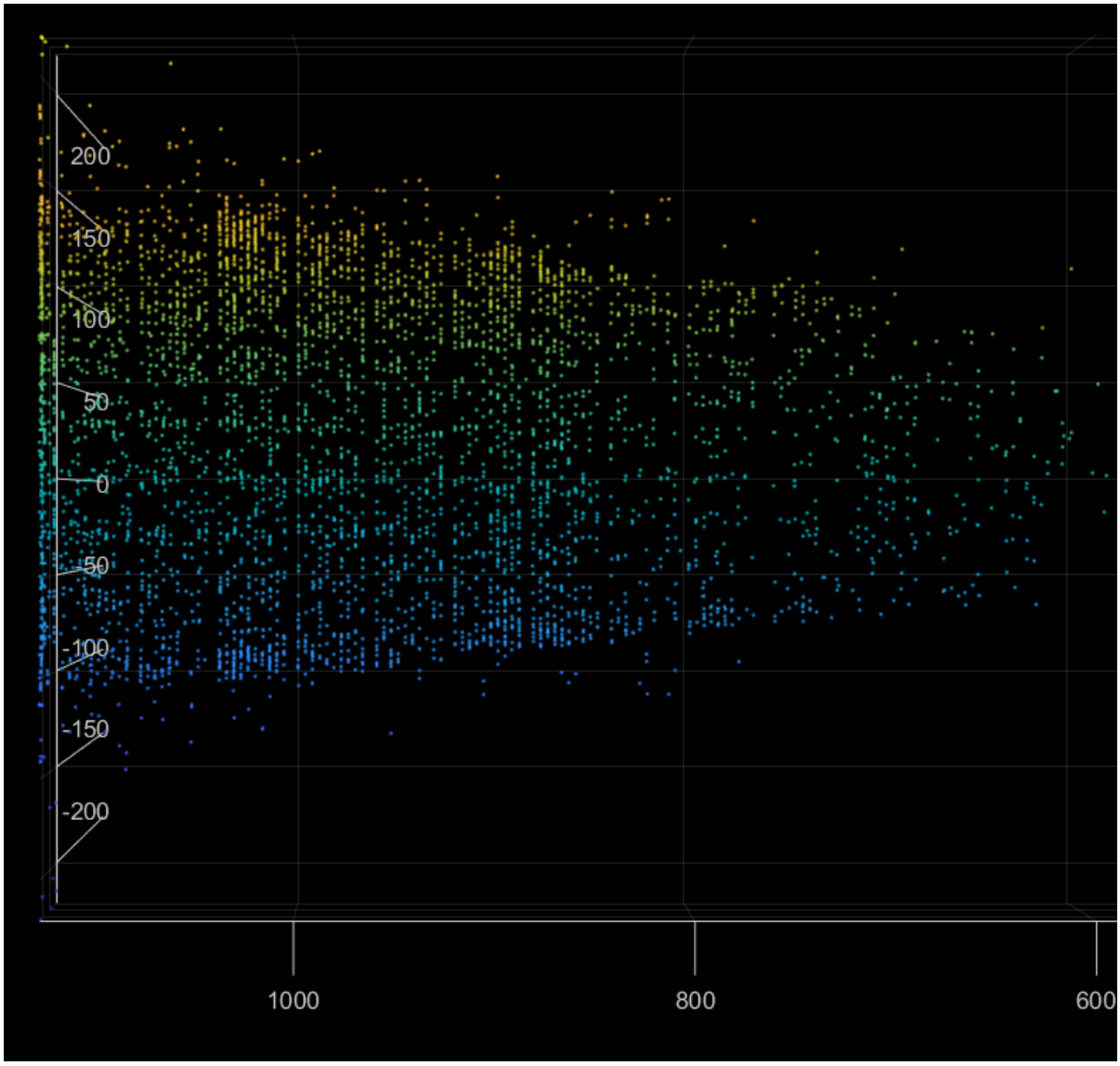


Figure 30: A top-down view of the point cloud seen in Figure 29

Discussion

The discontinuities seen in Figure 30 are consistently 32 mm apart; the squares in the checkerboard pattern are 25 mm x 25 mm indicating a consistent 7 mm error in the edge detection of the LoG algorithm. Solving for the equation of the line in , a change in 0.01V, the smallest increment of change possible with the liquid lens, corresponds to a change in d_o of 3.57 mm indicating that a 7mm error could be the result of the incorrect binning over the course of two focal power changes.

`

The liquid lens is significantly better at resolving the depth of the environment as it is capable of changing the voltage applied to the liquid lens, and thus changing the optical power, to a much finer degree than the geared lens setup was able to resolve the distance between the camera lens and the imaging plane. This can be attributed to the fact that there was backlash in the geared setup and the physical limitations of the stepper motor's step size. The liquid lens was driven by a voltage driver capable of manipulating the voltage by 0.01V corresponding to an optical power change of 0.0797, indicating that very small changes in optical power and d_o distances are achievable with the lens. This may have proved to be more detrimental than beneficial as it yielded overlapping depths of field.

The control of the optical power of the liquid lens was fine enough to have different regions of the environment in focus at different optical powers and thus create a depth map. The depth map in Figure 28, however, is subject to a significant amount of noise. The sharpness of the edges in the foreground persists throughout nearly all images in the focal stack; this likely indicates that an aperture of greater size is necessary to reduce the DoF. In addition to this, an objective lens such as the one used in Martel et al. (2018), would increase the utility of the liquid lens setup and allow it to resolve depths at greater distances by increasing the focal length.

Future Works

In future works, work should begin with a liquid-tunable lens and greater effort should be dedicated to selection of an objective lens for the liquid-tunable lens. Greater effort in this design area has a strong impact on the depth of field of the overall system and increases the depth resolution of the system. For example, when solving for the DoF of the current liquid lens setup, at a working distance of 500 mm and assuming the focal length of the lens is fixed at 12 mm, or the liquid lens is adding zero additional optical power, the DOF is approximately 150mm.

`

However, if the 35mm, f/7, Liquid Lens Cx Series Fixed Focal Length Lens from Edmund Optics is used with the same assumptions, the DoF falls to 13.67mm due to its increased focal length and aperture size. Table 2 has been compiled to show the DoFs corresponding to several liquid lens kits offered from Edmund Optics assuming an object of interest at a working distance of 500mm using Equation 2.

Table 2: A table of DoFs of Various Liquid Lenses offered by Edmund Optics

Name	f [mm]	d_i[mm]	$Aperture$ [mm]	C [mm]	DoF [mm]
12mm, f/6, Liquid Lens Cx Series Fixed Focal Length Lens	12	12.29508	2	0.007	145.2765
35mm, f/7, Liquid Lens Cx Series Fixed Focal Length Lens	35	37.63441	5	0.007	18.60644
50mm, f/7, Liquid Lens Cx Series Fixed Focal Length Lens	55	61.79775	7.857143	0.007	7.208639

Additionally, in future works, greater consideration must be given to a thresholding method or sampling function for the response to the LoG algorithm to reduce the noise captured in the depth map.

Conclusion

This work has examined depth-from-focus techniques presented in other works (Martel et al., 2018) to determine their efficacy on commercial-grade photography equipment. An initial proof-of-concept was presented that demonstrates the techniques are sound and commercial-grade equipment is capable of generating a depth map. This work has also identified that manipulation

`

of focus-tunable, electronically controller lenses are likely one of the only viable hardware-based methods to generate a depth map from a monocular source. This work has validated that the LoG algorithm is a reliable method of detecting edges as well as determining image clarity. This was useful in detecting in-focus pixels in focal stacks. Furthermore, a depth map was generated from the focal stack collected using the focus-tunable lens. Finally, this work identifies that greater consideration for the objective lens and aperture size used with the focus-tunable lens are critical for future works. A better sampling function is also required for reducing noise in the depth map. While the depth maps generated in this work were not as error-free as the depth maps produced in Martel et al. (2018), the proof-of-concept shows significant promise for depth-from-focus from commercial-grade photography equipment.

`

References

Bhairannawar, S. S. (2018). Chapter 4 - Efficient Medical Image Enhancement Technique Using Transform HSV Space and Adaptive Histogram Equalization. In S. S. Bhairannawar, *Soft Computing Based Medical Image Analysis* (pp. 51-60). Academic Press.

Bhuiyan, A.-A., & Khan, A. R. (2015). Image Quality Assessment Employing RMS Contrast and Histogram Similarity. *The International Arab Journal of Information Technology, Vol. 15, No. 6*, 983-989.

Bleser, G., Wuest, H., & Stricker, D. (2006). Online camera pose estimation in partially known and dynamic scenes. *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality* (pp. 56-65). Santa Barbara: IEEE.

Brown, M. (n.d.). *Mechanical vs electronic shutters*. Retrieved from Photo Review: https://www.photoreview.com.au/tips/shooting/mechanical-vs-electronic-shutters/

Cromie, G. (n.d.). *GUIDE TO THE CAMERA SHUTTER (FUNCTION, TYPES & MORE)*. Retrieved from Shotkit | Inspirational Photography & Free Guides: https://shotkit.com/camera-shutter/

Cruz, L., Lucio, D., & Luiz, V. (2012). Kinect and RGBD Images: Challenges and Applications. *25th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials* (pp. 36-49). Ouro Preto, Brazil: IEEE.

Dubey, A. (2020). *Stereo vision— Facing the challenges and seeing the opportunities for ADAS applications.* Texas Instruments.

Edmund Optics. (n.d.). *12mm Cx Series Fixed Focal Length Lens.* Retrieved from Optics Manufacturer & Supplier | Imaging Lens & Laser Optics: https://www.edmundoptics.com/p/12mm-cx-series-fixed-focal-length-lens/3229/

Edmund Optics. (n.d.). *Focus Tunable Lenses*. Retrieved from Optics Manufacturer & Supplier | Imaging Lens & Laser Optics: https://www.edmundoptics.com/c/focus-tunable-lenses/1418/

Etienne-Cummings, R., Kalayjian, Z. K., & Cai, D. (2001). A programmable focal-plane MIMD image processor chip. *IEEE Journal of Solid-State Circuits, vol. 36, no. 1*, 64-73.

ExpertPhotography. (2023, September 21). *Camera Lens Guide (Parts, Functions and Types Explained)*. Retrieved from ExpertPhotography Home - Courses and Traning for Photographers: https://expertphotography.com/camera-lenses-guide/

Fanani, N., Stürck, A., Barnada, M., & Mester, R. (2017). Multimodal scale estimation for monocular visual odometry. *2017 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1714-1721). Los Angelos: IEEE.

Fisher, R., S, P., Walker, A., & Wolfart, E. (2003). *Laplacian/Laplacian of Gaussian*. Retrieved from Digital Filters: https://homepages.inf.ed.ac.uk/rbf/HIPR2/log.htm

`

Flyps. (2023, January 4). *Depth perception in robotics (RGB-D). VSLAM use-case included*. Retrieved from Flyps - High-Tech Software House: https://www.flyps.io/blog/step-by-step-towards-the-depth-introduction-to-distance-perception-in-robotics#:~:text=Depth%20perception%20is%20the%20core,or%20more%20general%20scene%20understanding.

Fossum, E. R. (1989). Architectures For Focal Plane Image Processing. *Optical Engineering, vol. 28, issue 8*.

FRAMOS. (2023, April 6). *ADVANTAGES AND DISADVANTAGES OF TIME-OF-FLIGHT CAMERAS*. Retrieved from FRAMOS- See the Difference: https://www.framos.com/en/articles/advantages-and-disadvantages-of-time-of-flight-cameras#:~:text=ToF%20cameras%20do%20have%20some,confused%20by%20highly%20reflective%20surfaces

Geng, J. (2011). Structured-light 3D surface imaging: a tutorial. *Advances in Optics and Photonics Vol. 3, Issue 2*, 128-160.

Hawkins, C. (2017, November 20). *SIMPLE FORMULAS FOR THE TELESCOPE OWNER*. Retrieved from Sky & Telescope: https://skyandtelescope.org/observing/stargazers-corner/simple-formulas-for-the-telescope-owner/

Hesai Technology. (2023, May 25). *What You Need to Know About Lidar: The Strengths and Limitations of Camera, Radar, and Lidar*. Retrieved from Global Leader in LliDAR Sensor Solutions | HESAI Technology: https://www.hesaitech.com/what-you-need-to-know-about-lidar-the-strengths-and-limitations-of-camera-radar-and-lidar/

Hong, E., & Lim, J. (2018). Visual-Inertial Odometry with Robust Initialization and Online Scale Estimation. *Sensors vol. 18, issue 12*.

Kiatronics. (n.d.). *STEPD-01*. Retrieved from Electronic Components Distributor - Mouser Electronics: https://www.mouser.com/ProductDetail/OSEPP-Electronics/STEPD-01?qs=wNBL%252BABd93ONFTIfWMxSZw%3D%3D

Klein, G., & Murray, D. (2007). Parallel Tracking and Mapping for Small AR Workspaces. *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality* (pp. 225-234). Nara, Japan: IEEE.

Li, Y., & Ibanez-Guzman, J. (2020). Lidar for Autonomous Driving: The Principles, Challenges, and Trends for Automotive Lidar and Perception Systems. *IEEE Signal Processing Magazine, vol. 37, no. 4*, 50-61.

Martel, J. P., Müller, L. K., Carey, S. J., Müller, J., Sandamirskaya, Y., & Dudek, P. (2018). Real-Time Depth From Focus on a Programmable Focal Plane Processor. *IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 65, no. 3*, 925-934.

`

Mateer, P. (n.d.). *UNDERSTANDING THE CIRCLE OF CONFUSION IN PHOTOGRAPHY*. Retrieved from Shotkit | Inspirational Photography and Free Guides: https://shotkit.com/circle-of-confusion/

METTATEC. (2023, April 24). *What is LiDAR Technology? Advantages and Disadvantages*. Retrieved from METTATEC - GNSS Devices for Surveying and Photogrammetry: https://mettatec.com/what-is-lidar-technology-advantages-and-disadvantages/#Advantages_and_Disadvantages_of_LiDAR_Technology

Nagahara, H., Kuthirummal, S., Zhou, C., & Nayar, S. K. (2011). Flexible Depth of Field Photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 1*, 58-71.

NASA AMES Research Center. (n.d.). *LUMINANCE CONTRAST*. Retrieved from Color Usage Site: https://colorusage.arc.nasa.gov/luminance_cont.php

NASA. (n.d.). *Mars 2020 Perseverance Rover*. Retrieved from The Cameras on the Mars 2020 Perseverance Rover: https://mars.nasa.gov/mars2020/spacecraft/rover/cameras/

Nicholson, A., & Summersby, A. (n.d.). *Electronic shutter vs mechanical shutter*. Retrieved from Digital Cameras, Lenses, Camcorders & Printers - Canon Europe: https://www.canon-europe.com/pro/infobank/electronic-vs-mechanical-shutter/

Nikon. (n.d.). *Understanding Maximum Aperture*. Retrieved from Nikon | Shop & Explore Cameras, Lenses, and Accessories: https://www.nikonusa.com/en/learn-and-explore/a/tips-and-techniques/understanding-maximum-aperture.html

Nixon, M., & Aguado, A. (2012). Chapter 4: Low-level feature extraction (including edge detection). In M. Nixon, & A. Aguado, *Feature Extraction and Image Processing for Computer Vision* (pp. 161-166). Elsevier Science & Technology.

Nixon, M., & Aguado, A. (2012). Chapter 1: Introduction. In M. Nixon, & A. Aguado, *Feature Extraction and Image Processing for Computer Vision* (p. 8). Elsevier Science & Technology.

Nixon, M., & Aguado, A. (2012). Chapter 2: Images, sampling, and frequency domain processing. In M. Nixon, & A. Aguado, *Feature Extraction and Image Processing for Computer Vision* (p. 40). Elsevier Science & Technology.

Nützi, G., Weiss, S., Scaramuzza, D., & Siegwart, R. (2011). Fusion of IMU and Vision for Absolute Scale Estimation in Monocular SLAM. *Journal of Intelligent & Robotic Systems*, 287-299.

Onah, C. I., & Ogudo, C. M. (2014). Design and construction of a refracting telescope . *International Journal of Astrophysics and Space Science, vol. 2, issue. 4*, 56-70.

OpenCV. (n.d.). *Laplace Operator*. Retrieved from OpenCV: Opencv modules: https://docs.opencv.org/3.4/d5/db5/tutorial_laplace_operator.html

`

Optotune Switzerland AG. (2023, February 11). *EL-16-40-TC - Optotune*. Retrieved from Optotune: https://www.optotune.com/el-16-40-tc-lens

Parkhi, O. M., Vedaldi, A., Zisserman, A., & Jawahar, C. V. (n.d.). *The Oxford-IIIT Pet Dataset*. Retrieved from Visual Odometry Group - Oxford University: https://www.robots.ox.ac.uk/~vgg/data/pets/

Paul, J. (2016, July 25). *Rolling Shutter vs Global Shutter: What's the difference?* Retrieved from The Beat: A Blog by PremiumBeat: https://www.premiumbeat.com/blog/know-the-basics-of-global-shutter-vs-rolling-shutter/

Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America A, vol. 7, no. 10*, 2032-2040.

Photonics Media. (n.d.). *Gaussian and Newtonian Thin Lens Formulas*. Retrieved from Photonics.com: Optics, Lasers, Imaging & Fiber Information: https://www.photonics.com/Articles/Gaussian_and_Newtonian_Thin_Lens_Formulas/a25475

Sharifi, M., Fathy, M., & Mahmoudi, M. T. (2002). A classified and comparative study of edge detection algorithms. *International Conference on Information Technology: Coding and Computing* (pp. 117-120). Las Vegas: IEEE.

Szczys, M. (2012, May 6). *CONVERTING A MANUAL CAMERA LENS TO USE MOTORIZED ZOOM AND FOCUS*. Retrieved from Hackaday | Fresh Hacks Every Day: https://hackaday.com/2012/05/06/converting-a-manual-camera-lens-to-use-motorized-zoom-and-focus/

Teledyne FLIR. (n.d.). *Focal Plane Arrays*. Retrieved from Thermal Imaging, Night Vision, and Infrared Camera Systems | Teledyne FLIR: https://www.flir.com/browse/oem-cameras-components-and-lasers/focal-plane-arrays/

VectorNav. (n.d.). *VN-300*. Retrieved from VectorNav Technologies: https://www.vectornav.com/products/detail/vn-300

Vincent, K. (2023, July 2023). *An in-depth comparison of LiDAR, Cameras, and Radars' technology*. Retrieved from Outsight - unlocking the power of 3D LiDAR data: https://www.outsight.ai/insights/how-does-lidar-compares-to-cameras-and-radars

Vision Doctor. (n.d.). *Image sharpness and depth of field (DOF)*. Retrieved from Vision-Doctor - Industrial machine vision: https://www.vision-doctor.com/en/optical-basics/depth-of-field.html

Wang, Z., Shen, M., & Chen, Q. (2023). Eliminating Scale Ambiguity of Unsupervised Monocular Visual Odometry. *Neural Processing Letters, 55*, 9743–9764.

`