NORTHWESTERN
UNIVERSITY

Computer Science Department

Technical Report
Number: NU-CS-2023-07
Februrary 2023

SPIDER INSPIRED MOBILE CAMERA MODEL DESIGN

Yiran Zhang

Abstract

We will configure three families of hunting spiders, crab spiders, wolf spiders, and jumping spiders, and see how they achieve depth information differently. Meanwhile, their multiple eyes design working along with their different depth estimation methods will also proven to work for recent mobile photography. This paper also includes the correlation between spider vision and this perspective camera design model.

Keyword

Depth from Defocus, Stereo, Depth Sensing, Range Imaging, Mobile Photography, 3D Vision, Bio-inspired Vision.

ABSTRACT

Some animals in nature survive by receiving depth information in scenes through different approaches, like hunting spiders. We will configure three families of hunting spiders, crab spiders, wolf spiders, and jumping spiders, and see how they achieve depth information differently. Meanwhile, the unique visual system of jumping spiders pushes this work to come up with a depth estimation method that also proves to work for recent mobile photography. Due to more cameras being added to our phones, this paper presents a training model that can measure desired camera arrangement designs by combining Depth from Differential Defocus (DFDD) and stereo methods. The result camera design can well estimate the depth information in a wide depth range. This paper also includes the correlation between spider vision and this perspective camera design model.

ACKNOWLEDGEMENTS

I really would like to thank Dr. Emma Alexander for advising me on this thesis and Dr. Jack Tumblin for being on my committee. I am truly grateful for all their inspiration, patience, and understanding of them throughout the whole process. As a kid who just stepped into 3D vision, I will not be able to present this work without their help. Their passion for knowledge and academics also motivates me to ”rise above and focus on science”, as Grandpa Rick said. I also appreciate the help from Deniz Korman and his kind offering on spider images in this paper.

Moreover, I want to thank my parents for supporting me and my decisions all the time. Their love and encouragement are always the main energy that fills me whenever frustration comes. I also really appreciate my grandma's understanding of the pursuit of my goal. Thanks for her always telling me about her trust in me and for being a person just like I expected from myself.

Finally, I am really grateful for the whole thesis experience where I learned about not just related fields and mathematics behind different depth estimation methods, but also more about myself. In the process, I begin to realize my characteristics' advantages and disadvantages, where I need to train myself more, and when I need external help. I start to understand and build up my intention to research during my master's study at Northwestern, find my own pace of active learning, and elevate my level of understanding to knowledge. I believe this new building-up realization and acknowledgment of myself will continue polishing me into a mature researcher and empower me for pursuing my research goal.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

A large number of creatures in wild have their own way of viewing or feeling this world completely different from us humans. Some animals, like urchins, have a low-resolution vision but have special light-sensitive cells on their feet that help them to ”locate” themselves [1]. Some animals heavily depend on their 2D vision to recognize and survive in the wild. Also, some creatures, like humans, are able to receive depth information in the scene. Hunting spiders are one of the species whose survival and behaviors heavily depend on depth recognition. With their several pairs of eyes, they are able to use both stereopsis and focusing or defocusing methods to detect the surrounding objects or things far away from them.

Their ability to achieve depth information from a scene to construct the 3D space thus makes a lot of researchers interested and eager to understand their vision system better. Among all the interesting families of hunting spiders, this work focuses on the stereopsis and focusing system of Thomisidae, crab spiders, and Lycosidae, wolf spiders, and the stereopsis and defocusing system of Salticidae, jumping spiders.

The specialties of the jumping spider also inspired this work to push the depth estimation to the current mobile photography. Similar to the hunting spiders' multiple pairs of eyes, most current mobile phones implement more than one camera. Multiple cameras can computationally implement stereo, focusing, or defocusing approaches in order to achieve more accurate depth results. They can also offer different aperture sizes for shooting scenes from a distinct depth of field. This presented model uses the stereo and depth from differential defocus (DFDD) methods [2] and calculates the best camera arrangement of two lenses so that the depth estimation can work together to reach the best depth result for a wide depth range.

CHAPTER 2

RELATED WORK

2.1 Biological Vision to Computational Vision

Depth information is essential for some creatures of animals and thus different ways of collecting depth information vary from each other. Human uses stereopsis and focusing method to achieve depth [3]. In nature, animals use their unique way of understanding the world, and most biological depth perception strategies fall into four categories as 2.1 shows. The monocular and multiocular systems by the number of sources. Static or dynamic identifies whether the animal's vision is instantaneous or it detects changes over time. Among all four kinds, hunting spiders are believed to have static multiocular vision. Due to the slight changes in focal length between each pair of eyes, their vision is a long-term biological research topic.

Besides spiders, various kinds of animals have unique own advantages in their vision systems. Biological vision systems have always been a great source of inspiration for various computer vision algorithms and new camera designs. Bees have a higher sensitivity to spectrum up to the ultra-violet range and a much wider field of view than human vision [4]. Although not able to see the ultraviolet spectrum through our eyes, a 160 degrees field of view camera, the arthropod camera, was designed in 2013 to mimic bees to achieve a vision system [5]. Also, helmet geckos have extremely good night vision and are believed to have a multi-focal optical system where they can adjust their "aperture size" depending on the brightness of the scene [6]. A gecko-inspired sensor was created under the inspiration of nocturnal geckos' color vision under dim light. Two image intensifiers were included in this sensor to act like night vision goggles [7].

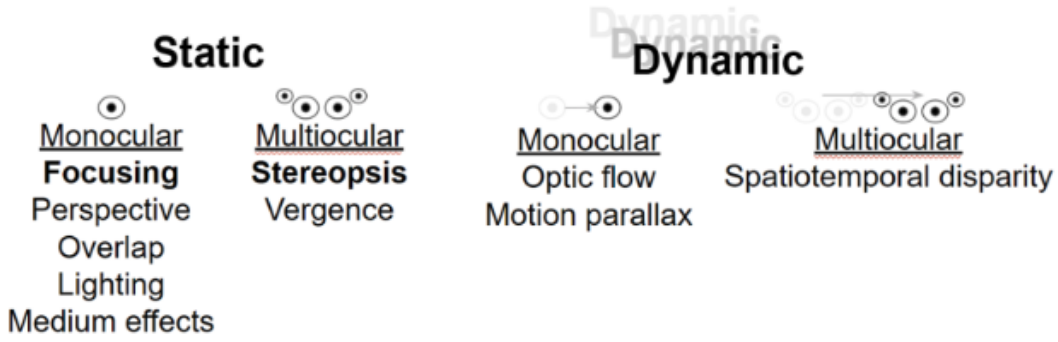

Figure 2.1: Animal Eye Types in Nature (Figure Credit: Deniz Korman)

2.2 Depth from Defocus vs. Stereo: How Different Really Are They? [8]

In this paper, Schechner and Kiryati [8] mentioned the commonness and the differences between Depth from Focus and Defocus and stereo or Depth from Motion Blur. They conclude the previous beliefs about those methods: 1. the sensitivity of Depth from Focus and Defocus (DFF, DFD) methods does not behave as well as the triangulation-based techniques; 2. DFD avoids the occlusion problem more than the stereo technique; 3. DFD avoids matching problems which the stereo approach always suffers on; 4. DFD is more reliable when responding to perturbation.

Further in his paper, the first three were proved wrong and the fourth one is still sensible. The stereo sensitivity is related to the smallest detectable disparity. For the sensitivity analysis, they proved in the case of DFD, it can be treated the same as two stereo pinholes separated by a baseline D as the DFD aperture size (**??**). The sensitivity of both methods is the same when relating the stereo's disparity to the PSF support size. Moreover, the sensitivity of DFF is also similar to stereo because the disparity of the stereo method and the blur diameter are sensed similarly.

For the occlusion problem, they show that either DFD or DFF can ensure avoiding this problem. DFD was proved to suffer this problem more than the stereo approach. The chief array of DFD is larger than stereo and thus the possibility of the blur PSF being covered is also higher. In the DFF method, they split the chief ray image points into four classes and analyzed each with the same occluder position. If the occlusion is small, DFF yields a good depth value. On the other hand, if the occlusion is severe, both methods yield ambiguous results while depth convergence yields

correct depth estimation.

The matching ambiguity also proved to not just exists in stereo but also in the DFD method. Stereo needs the detection of disparity between the left and right images of a common object to measure depth. Matching distinct images together thus rises a matching problem. In the DFD method, depth is measured by matching the same area through image patches through different focuses. In the aperture changing DFD method and Depth from Motion Blur (DFMB) method, the difficulty of matching issues exists for the transfer function between the DFD images. Schechner and Kiryati believe that only one-to-one transfer function models can resolve this problem. On the other hand, the focus-changing DFD method cannot avoid this problem due to the same reason as the non-one-to-one transfer function. Nevertheless, a special case when blur diameter changes $\nabla d \rightarrow 0$ can ignore the matching ambiguity issue.

Moreover, for the same system dimensions, common triangulation techniques may be less prone to matching ambiguity than DFD. Unlike DFD and stereo, DFF seems true to be immune to matching ambiguities, if the evaluation patch of the focus measure is larger than the support of the widest blur kernel expected, and if the depth is homogeneous in that patch. The fourth statement on managing the perturbation makes DFD and DFF a more solid method based on a single frequency perturbation. Since they rely on more data than the triangulation approach, it is reasonable to be more reliable.

2.3 A Theory of Depth from Differential Defocus [2]

Depth From Differential Defocus (DFDD) visual system is inspired by jumping spiders' vision which has a stack of four retinas, each at a slightly different scale of blurriness to help them observe the small change in the environment through the defocus level. In this paper, Alexander recommended a new type of sensor that uses the defocus brightness constancy constraint for estimating depth from small optical defocus changes to achieve fewer error metrics on the estimation result. Many differential changes to camera parameters can cause a defocus change that reveals depth through the defocus brightness constancy constraint. Parameters include a sensor location,

an aperture width, an optical power, a lens location, a camera location, and a combination of factors. Defocus brightness constancy constraint is a way of using brightness constancy by measuring the changes in blur kernels. Each pixel of the image change is able to be described as a weighted sum of spatial image derivatives. Finally, the depth of each pixel can be calculated through the image derivative (generated from the image change equation) and the blur scale values by plugging in the parameters of each specific case. The depth detection function and error metric that this presented model used will be later explained in section 5.3.

CHAPTER 3

SPIDER VISION

Hunting spiders (figure 3.2), such as jumping spiders (Salticidae), wolf spiders (Lycosidae), and crab spiders (Thomisidae), all have four pairs of eyes: posterior lateral, posterior median, anterior lateral, and anterior median (figure 3.2). In particular, the vision system of hunting spiders allows them to navigate 3D environments, identify court mates, and catch prey from a distance [9, 10]. Their different pairs of eyes also function differently with each unique focal length and acuity [11]. Some have a wider field of view than others, some can use the vergence technique to change both eyes' angles in order to gaze at one target, some baseline disparities might contribute to stereopsis, and some monocular/multiocular eyes might obtain depth information through focusing [11].

Specifically, jumping spiders are proven to have multiple tiers which means they have several layers of retinas in their monocular eye(s) [12]. With a slight difference in the DoF between their tiered retinas, their monocular eyes are able to compare the small changes of blurriness between retinas. Therefore, they can gain depth estimation of their targets out of the measurement of the different levels of defocusing between tiers [13]. The jumping spider is hypothesized to be the only animal that used monocular depth from defocus strategy to receive depth information.

Moreover, each pair of jumping spider eyes shows a linear chromatic aberration, a field of views, a unique wavelength of light, and a distinct focal distance [13, 11]. Especially, the distinct focal distance of their eyes offers them the ability to better detect the surroundings in a range that is much further than their own body size.

Animals who use Depth from Focus can detect the sharp vision of a specific depth similar to humans [3]. Comparably, Depth from Defocus allows animals to view the same object with different levels of focus and use the blur differential through their four tiers retinas to estimate the depth of the target. For instance, jumping spiders' defocus change can cause a change in the magnification of the underlying image [11]. This differential defocusing computation thus inspired

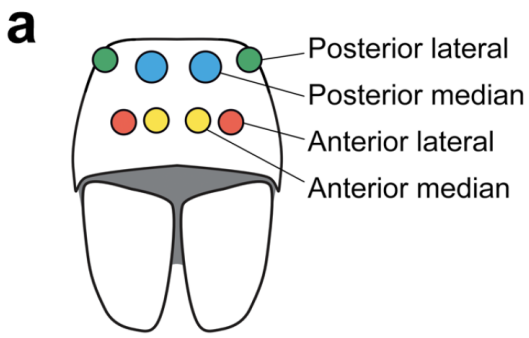

Figure 3.1: The arrangement of Spider Eye Pairs (Figure Credit: Deniz Korman)

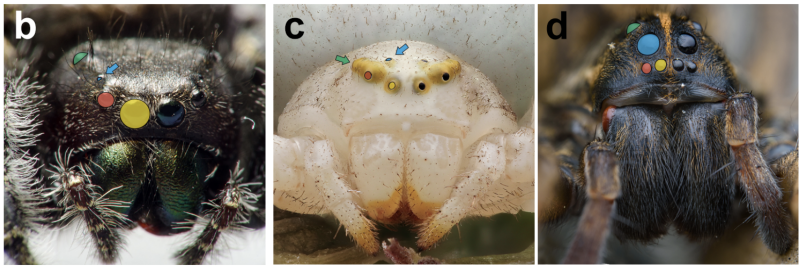

Figure 3.2: jumping spiders (Salticidae), crab spiders (Thomisidae), and wolf spiders (Lycosidae) (Photo Credit: (b) Colin Carmichael, (c) Andrius Kybartas, (d) Dalton Ramsey, and Color Matching on Photo Credit: Deniz Korman)

the computational vision field.

Relating to vision technology, an additional algorithm for extracting scene depth by comparing the image values between and across the retinal layer can is promoted to work efficiently with a spider-plausible image formation model [2]. The resulting image can be modeled by measuring the slight changes between the retinal layers and the differential changes across every single layer. The depth under the back projection of photoreceptors will be calculated by the derivatives of those changing variables. Nevertheless, it would be hard for the current market to mount a stack of photosensors like the jumping spider retinal layers at a reasonable price.

CHAPTER 4

STEREO VISION

4.1 Stereo Photography

In the 3D vision, triangulation is a common way of identifying an object in 3D space. Contrary to monoculars, the stereo vision was initially implemented in camera photography and originally used two or more lenses to take separate photos to map the difference in object position in between [14]. In current mobile photography, multiple cameras are implemented on phones in order to get a system that receives slightly different angles of the same objects. Several cameras are set at different locations to form a camera system and take multiple images at the same time. By analyzing the matching points shift in each image, the depth of the target objects can be calculated [15].

LG Optimas 3D, iPhone 8, and a lot of our current phones that are propagating their photography performance have multiple cameras implemented that act for stereo triangulation [15, 16, 17]. The implementation of a passive stereo vision system in mobile started with HTC EVO 3D. Stereo photos, videos, and displays were all allowed in this 2011 phone. This was also considered to be one of the mobile platforms that are available for 3D models and games [18].

Stereo vision can be divided into passive and active stereos. Passive stereo sensors like TaraXL [19] can thus work well in a good lighting scene. However, when the scene does not have a well-constructed light, this system is not able to do accurate boundary and shape detections [20]. On the contrary, the active stereo system uses a structured light, normally a random pattern projector (RPP) between the cameras, to obtain better images for the later detection process. Due to the less requirement for scene lighting, this system is considered to be more adaptable in 3D mapping [21]. Nevertheless, this system also limits its depth range of the detecting scene due to the natural lighting loss [22].

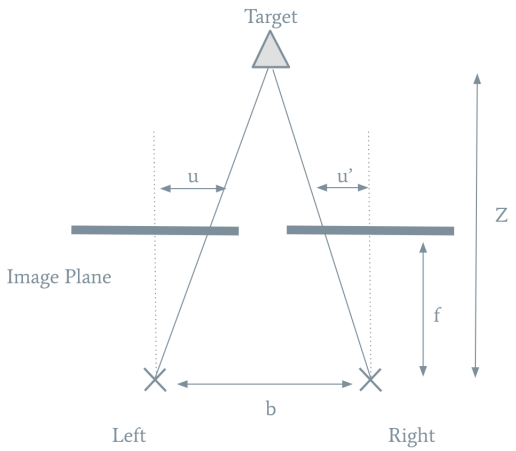

Figure 4.1: Naive Stereo System. (Z: object depth; b: stereo baseline; f: focal length; u: object shift in the left image plane; u': object shift in the right image plane)

Recently, the stereo strategy is normally used along with various kinds of machine learning algorithms to train and receive better performance [23, 24, 25, 26].

4.2 Mathematical Equation

According to Schechner and Kiryati [8], we here would like to view the stereo matching problem as a solved problem and further explain the lower bound of the stereo error metrics. The stereo approach is a triangulation method that estimates depth based on the distance between both cameras. In fig 4.1, we assume both stereo cameras are the pinhole cameras and are thus all in focus. The object distance is assigned as Z. The stereo baseline distance between the left and right camera. An estimated disparity $\Delta x = u + u'$ between a target point in the left image and the right image.

Then, it is obvious to see the similar triangles that are constructed from the target to the image planes and from the target to the left and right cameras. Therefore, we can get an equation

$$\frac{Z-f}{Z} = \frac{b-\Delta x}{b}. \tag{4.1}$$

Therefore, the depth Z can be solved as

$$Z = \frac{bf}{\Delta x}. \tag{4.2}$$

Without considering the correspondence problem mentioned in section 2.2, our assumption is the two images are perfectly matched. While measuring the sensitivity of the stereo approach, the error on u and u' occurs because a more complex stereo system may have a separate algorithm to calculate the image disparities corresponding to different image plane designs or camera system designs. Therefore, the sensitivity can be measured with respect to Δx. The stereo error metrics can be formulated as its first derivative of depth Z

$$|\frac{dZ}{d(\Delta x)}| = |\frac{bf}{-(\Delta x)^2}| = \frac{Z^2}{bf} \tag{4.3}$$

whereas the resulted lower bound error is a scalar that can be multiplied to the disparity detection with different stereo system camera models.

CHAPTER 5

DEPTH FROM DEFOCUS AND DIFFERENTIAL DEFOCUS

5.1 Aperture and Depth of Field

In photography, focusing on a specific expected depth range is important and there are a lot of interesting related notions related to the focusing technique. One of them that I mentioned several times in the introduction section is called depth of field (DoF). The distance and value of DoF can be changed with different aperture sizes. To represent aperture sizes for lenses, an f-number whereas f represents the focal length is needed. When reading the camera information, the f-numbers N are written as

$$N = f/D \tag{5.1}$$

where f is the focal length of the lens and D is the entrance pupil of the lens. $D = A * 2$ where A is the aperture radius of the given lens Therefore, this means the larger number of entrance pupils D is, the smaller aperture it gets, and a deeper depth of field photo can be provided so that a good great amount of in-focus details in all distances can be caught. Comparably, a smaller D value, a wider aperture, and a shallower depth of field image can be created to offer a good catch of details and lighting in only the expected area and blurriness in the rest.

5.2 Depth from Defocus

Depth from Defocus (DFD) takes each image patch in an image and measures the level of blurriness at different depths to calculate the depth. The degree of defocus of each image patch is thus based on the scene focus and the defocus Point Spread Function (PSF). Analyzing the blur, the focal point is further away than the image sensor from the aperture. Thus, we can get the location of the target object o from the distance from the image sensor to the aperture s, the blur circle size on the image sensor, the in-focus location, and the aperture size (Figure 5.1).

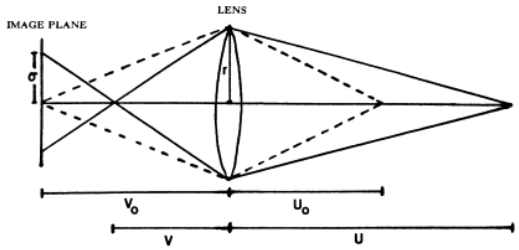

Figure 5.1: Depth from Defocus Camera Model (v_0: the distance between the image plane to the lens; u_0: the distance between the lens and focus point; when observing the target point in the distance of $u > u_0$, the point is in-focused at the distance of $v < v_0$, and thus a blur scale of σ is formed) (Figure credit: Alex Pentland)

Depth from Defocus (DFD) can measure the depth by analyzing the amount of defocus simultaneously at all points by taking one or two images. It is a depth estimation method first mentioned by Pentland [27]. He suggested using the point spread function (PSF) to measure the changing image intensity through different patches. The Fourier domain equation of the relation between the two blur circles' sizes in two images. An equation of object distance Z came up to be

$$Z = \frac{f v_0}{v_0 - f - \sigma N} \tag{5.2}$$

where f is the focal length, N is the f-number, v_0 is the image distance, and σ is the blur scale on the image plane shown in above figure 5.1 [27].

5.3 Depth from Differential Defocus

If a camera has the observation of the changes in the level of defocus, the defocus brightness constancy constraint can be generated. If we consider the DFDD method as a pair of stereo cameras in 2D, the block with the size of the stereo baseline is ignored. A DFDD sensor with the stereo baseline $D = A * 2$ as aperture diameter D and aperture radius A can produce the depth Z calculated from the distance from the image plane to the lens f, aperture size D, in-focus depth Z_f determined by f and lens' optical power p, and the estimated blurring radius $\tilde{A}$ [2]:

$$Z = \frac{Z_f f A}{Z_f \tilde{A} + f A}. \tag{5.3}$$

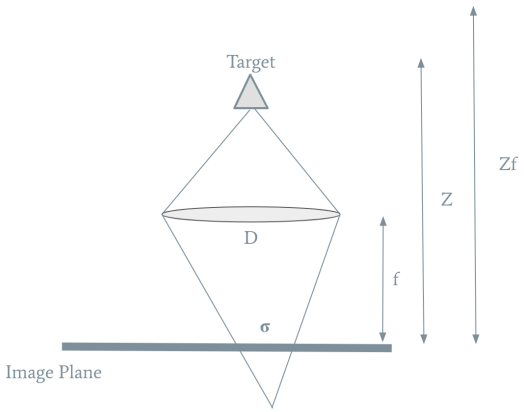

Figure 5.2: Naive DFDD System. (D: aperture diameter; Z: object depth; Zf: in focus depth; f: image plane distance; σ: blur scale)

As calculated for the DFDD sensitivity, the error in the variable $\tilde{A}$ on depth Z can produce the DFDD error metrics as [8]

$$|\frac{dZ}{d(\tilde{A})}| = |\frac{-Z_f^2 fA}{(Z_f\tilde{A}+fA)^2}| = \frac{Z^2}{Af}. \tag{5.4}$$

Recall the stereo sensitivity equation as 4.3, the 5.4 equation acts the same due to analogously considering baseline in stereo as a big aperture in defocus method [8]. However, stereo and DFDD do have completely different focal depths. In the real-life case, stereo consists of two pinhole cameras whereas defocus cameras have specific focus depth relative to the distinct aperture size as explained in section 5.1. The stereo equation 4.3 make sense since its parameters do not deal with the actual aperture in the camera system. Comparably, the aperture radius A is in the DFDD equation and thus we cannot simply ignore the focal depth of the lens.

As figure 5.2 shows, Zf represents the focal depth. As Alexander later illustrated in [2], the DFDD approach is sensitive to the distance between the camera and the image plane through the $|Z - Z_f|$ term. In our demonstrating model, we change the above Z^2 term into $Z|Z - Z_f|$ and simplified the DFDD error metrics into

$$\varepsilon \leq \frac{Z|Z - Z_f|}{Af} \tag{5.5}$$

whereas the resulted error is a also scalar value that can be multiplied to the blur circle detection

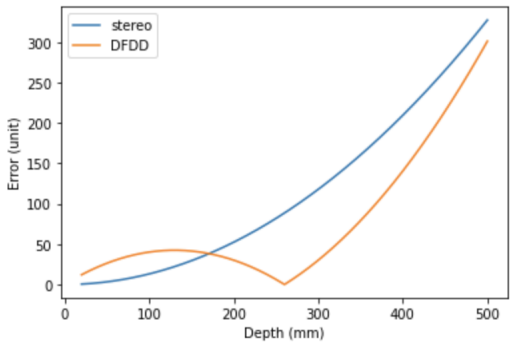

Figure 5.3: Naive Stereo and DFDD Model (the x-axis gives depth points in a range of depth (20mm,500mm); the y-axis shows the Stereo and DFDD's MSE at different depths; the blue line represents the stereo; the orange line represents the DFDD.)

with different defocusing camera models.

Therefore, the 5.5 equation is a more reasonable error function to utilize in a camera with the DFDD implementation. This equation makes more sense not only for the cameras but also for the monocular defocus jumping spider eye. The four tiers of a jumping spider's monocular eye have multiple levels of focusing through their different retina locations in the eye. Hence, although with the same aperture size (in a monocular eye), the Zf value of each tier varies. The error function shown in 5.5 can thus be able to detect this variation.

The finalized DFDD error function in 5.5 and the stereo error function in 4.3 can thus offer us an interesting conclusion of both methods in figure 5.3. In the depth range of (20mm to 500mm), the stereo method offers less error in near vision. However, as the depth goes larger, the DFDD appears to have less error than the stereo. This is because the numerator of DFDD equation 5.5 allows this method to detect small changes in depth around Zf like things in the far vision. However, the stereo approach is not able to achieve this due to its pinhole images. Therefore, our model will combine both methods and come up with a model to get the benefit of both 5.4.

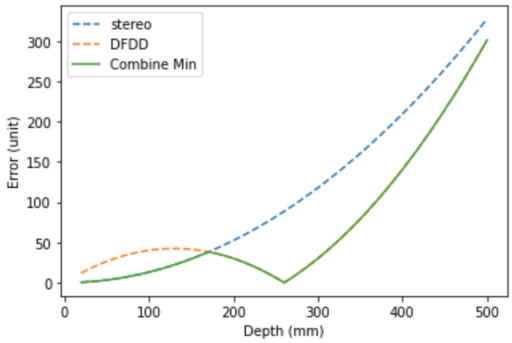

Figure 5.4: Naive Stereo and DFDD Model (the green line represents the minimum error of both methods at each depth point.)

CHAPTER 6
MODEL IMPLEMENTATION

6.1 Thinking Back to Spiders

Recall the vision system of stereopsis and focusing of wolf spiders, crab spiders, and defocusing of jumping spiders that we explained in chapter 3. In our assumption, the spider families are able to process the best depth estimation approach with their lens design. We observed that the wolf spider and crab spiders who are believed to use the all-in-focus stereopsis method have pairs of pinhole eyes in order to allow a larger disparity in both eyes. They also appear to use their eyes to hunt in near vision. Comparably, the jumping spiders who have big aperture eyes appear to hunt further and thus were believed to use their aperture as depth from defocusing technique to precisely detect the distance of prey in far vision.

6.2 Model Overview

As the vision system of wolf, crab, and jumping spiders, this model is eager to explore how the different visual system for those spiders works for their hunting behavior. Reflecting on the two different vision systems of spiders who hunt in the near and far field, this model also includes how the lens can be best arranged in the middle field while using the stereo and DFDD combined algorithm.

6.3 Initialization

Given the error metrics of both approaches, equation 4.3 for stereo and equation 5.5 for DFDD, the best arrangement of a pair of eyes to see the target in a specific depth range can be plotted as a graph like the green line in Figure 5.4. Distinctly, the jumping spiders' targets are normally in a smaller range than what Figure 5.4 shows. In order to best guess their arrangement of the eyes'

aperture size and stereo baseline length within each depth range, and how they trade off for both methods and finalize a lens arrangement with a long enough baseline and a big enough aperture to obtain the depth information, a gradient descent algorithm is implemented to optimize on the combined (stereo and DFDD) loss function. In this case, the result lens arrangement would be the most possible design of the spider visual system in the given depth.

The expected result of the model is to estimate the best arrangement of two lenses in a given area in order to present the desired image for an object in a specific depth range. The stereo and DFDD approach error metrics were stated in equations 4.3 and 5.5. The training parameters for both error metrics functions include the stereo baseline, the focal depth value, and the aperture radius.

For the sake of the variables' stability during the later gradient descent process, we initialized the training theta

```
thetas = np.array([random.uniform(0.01,0.1),
                   random.uniform(0.3,0.9),
                   random.uniform(0.01,0.1)])
```

with three parameters in sequence are: $theta\_b$, $theta\_Zf$, and $theta\_a$ which represent the ratio (from 0 to 1) of the maximum of each variable (stereo baseline, focusing depth, and right aperture size). The stereo baseline and right aperture theta start with small values and $theta\_Zf$ was assigned randomly between (0.3,0.9) to allow a fair initialized focusing depth within the depth range (for DFDD approach). With a fixed area with the length of max_length and the minimum aperture size r_min of both lenses, we can calculate the maximum baseline $max\_baseline$ for a range that totally wins on the stereo approach and a maximum aperture size max_r_dfdd when the depth range totally wins on the DFDD approach as:

```
max_baseline = max_length - 4 * r_min
max_r_dfdd = (max_length - 2 * r_min)/2.
```

Also, the focal depth Zf in equation 5.5 is also being trained within the range of each given depth

range as:

```
range_Zf = (max(Zs) - min(Zs))
min_Zf = min(Zs)
```

where *Zs* is a list of depth values extracted from the given depth ranges. The resulting float for each variable's theta value will multiply by each variable's maximum value. Therefore, the variables in each training echo can be represented as:

```
b = max_baseline * theta_b
Zf = range_Zf * theta_Zf + min_Zf
A = max_r_dfdd * theta_a
```

where b, Zf, A are the actual values of the stereo baseline, focal depth of both lenses, and the right aperture radius.

Finally, relating back to equations 4.3 and 5.5, the stereo and DFDD error metrics can thus be presented as:

```
err_stereo = (Zs * Zs)/(b * f_stereo)
err_dfdd = (Zs * abs(Zs - Zf))/(2 * A * f_dfdd)
```

where Zs is a list of depth values in the given depth range, and f_stereo and f_dfdd are the focal length values for both lenses. The *min_result* to calculate the minimum value of each depth value in *Zs* (similar to what Figure 5.4 shows)can be assigned as:

```
min_result = np.minimum(err_stereo ,err_dfdd).
```

6.4 ReLU Implementation

Since this lens arrangement was constrained by a *max_length* length, a ReLU function is implemented in order to constrain the racing between stereo baseline and aperture size. Rectified linear activation (ReLU) is a method to constrain one or more variables to a certain range [28]. In our model, we implement ReLU as a loss function in order to add a penalty when the

length_used > *max_length*. The *length_used* can be calculated and the ReLU loss function can also be implemented.

```
length_used = b + 2 * r_min + 2 * A
ReLU_constrain = ReLU(abs(length_used - max_length))
```

where the *ReLU_constrain* function restricts the variables inside the given *max_length*. If the *length_used* > *max_length*, *length_used* − *max_length* can be tuned to 0.

6.5 Implementation of Each Gradient Descent Epoch

In the gradient descent function, a function *loss_function* is defined and called for the given iteration number of training. From the last two sections, we can initialize the training variables as:

```
def loss_function(thetas, penalty):
    thetas[0] = theta_b
    thetas[1] = theta_Zf
    thetas[2] = theta_A
    b = max_baseline * theta_b
    Zf = range_Zf * theta_Zf + min_Zf
    A = max_r_dfdd * theta_a.
```

Both stereo and DFDD error metrics function, and the minimum function as:

```
    err_stereo = (Zs * Zs)/(b * f_stereo)
    err_dfdd = (Zs * abs(Zs - Zf))/(2 * A * f_dfdd)
    min_result = np.minimum(err_stereo, err_dfdd).
```

In order to do gradient descent, we can use the *GradientTape* which is a helpful function to calculate the derivatives of the complex combined loss function [29]. Firstly, we set the *theta_b*, *theta_Zf*, *theta_a* as the variables:

```
theta_b = tf.Variable(thetas[0])
theta_Zf = tf.Variable(thetas[1])
theta_a = tf.Variable(thetas[2])
```

The gradient tape can thus be implemented where the stereo function *stereo_loss*, the DFDD function *dfdd_loss*, and the ReLU constrain function *ReLU_constrain*.

```
with tf.GradientTape(persistent = True) as tape:
    stereo_loss = (Zs*Zs)/(max_baseline*theta_b*f_stereo)
    dfdd_loss = (Zs*abs(Zs-Zstart-Zf*theta_Zf))/
        (2*max_r_dfdd*theta_a*f_dfdd)
    length_used = theta_b*max_baseline +
        max_r_dfdd*2*theta_a+ 2*r_min
    ReLU_constrain = ReLU(abs(length_used - max_length))
```

Recall Figure 5.4, we want the combined function in each depth value in the given depth range to be able to reach as many minima of stereo and DFDD approach as possible. Hence, the combined MSE of the scalar value from error function *y* is constructed with the parts: the error *min_stereo* where stereo hit the *min_result* of the depth value, the error *min_dfdd* where DFDD hit the *min_result* of the depth value, and the ReLU constrain function *ReLU_constrain* to limit the training inside the given length range.

```
    min_stereo = (min_result == err_stereo).astype(int)*loss1
    min_dfdd = (min_result == err_dfdd).astype(int)*loss2
    y = tf.math.reduce_sum(tf.math.square
        (( min_stereo + min_dfdd + ReLU_constrain)))/len(Zs)
```

Then, we are able to assign the gradient variables:

```
grad_b= tape.gradient(y, theta_b)
grad_Zf = tape.gradient(y, theta_Zf)
grad_a= tape.gradient(y, theta_a)
```

For each training epoch, the gradient descent can be applied as:

```
opt.apply_gradients(zip([grad_b, grad_Zf, grad_a],
    [theta_b, theta_Zf, theta_a]))
```

Finally, the theta value of the three variables are able to be updated as:

```
thetas[0] = theta_b
thetas[1] = theta_Zf
thetas[2] = theta_a
```

After looping this process with a certain training rate in the given number of iterations, the three variables, stereo baseline, focal depth, and right aperture size, can be trained and are able to construct the best arrangement of two lenses to view the target at the given depth range.

CHAPTER 7

MODEL RESULT AND OTHER INSPIRATION

7.1 Result Demonstration

As the training process explained in the last chapter, we plot the final trained lens arrangement result as images with a *show_result*() function and analyze them in the case of spiders. The result includes an illustration of the visual system of the three spider families that we explain in chapter 3 and further exploration of a wider depth range.

Recall the visual system of crab spiders and wolf spiders explained in chapter 3. Both crab spiders and wolf spiders use stereopsis and focusing systems and match the images from both pinhole eyes. The first image of Figure 7.1 enables us to see how their anterior lateral eye pair which focusing and stereopsis vision system is preferred in the near-depth field, such as 20mm to 50mm. Within the blue area 50 mm of total length, the lens can be arranged as a big baseline of 46mm where both lenses' pinhole diameter is 2mm as the green circles/dots show. With the all-in-focus images that pinhole cameras of the left and right created, a larger disparity can help them to better identify the depth corresponding to the error metric in equation 4.3.

In Figures 7.1, within the blue area of total length, both green circles represent the two arranged lenses and are separated by the trained baseline between each other. It is obvious to see that when the distance is nearer, the algorithm recommends a larger baseline between the cameras along with a smaller aperture because the stereo approach strongly wins over. As the distance gets further, the model prepares a larger right aperture size for the DFDD algorithm due to its higher sensitivity to small changes (and thus moving objects in further vision field).

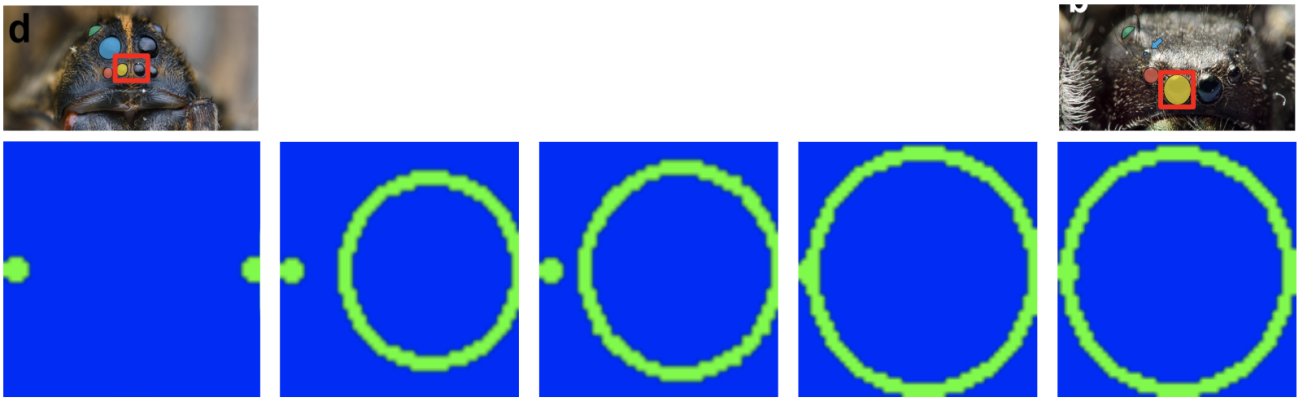

Figure 7.1: These are the camera/lens arrangement with combined (stereo and DFDD) algorithms in depth ranges: (20mm,50mm), (30mm,70mm), (80mm,120mm), (100mm,200mm), (300mm,500mm). Hunting in depth range of (20mm, 50mm): two pinhole eyes with an aperture size of 2mm, both with a focusing depth of 47.175 mm, and a stereo baseline of 46mm. Hunting in depth range of (30mm, 70mm): one pinhole eye and one eye with an aperture size of 36.267mm, both with a focusing depth of 53.071mm, and a stereo baseline of 11.734mm. Hunting in depth range of (80mm, 120mm): one pinhole eye and one eye with an aperture size of 40.779mm, both with a focusing depth of 102.259mm, and a stereo baseline of 7.218mm. Hunting in depth range of (100mm, 200mm): one pinhole eye and one eye with an aperture size of 47.0276452262mm, both with a focusing depth of 160.959mm, and a stereo baseline of 0.972mm. Hunting in depth range of (300mm, 500mm): one pinhole eye and one eye with an aperture size of 47.999mm, both with a focusing depth of 421.859mm, and a stereo baseline of 0.000mm.

7.2 Result Analysis

Recall the visual system of the jumping spider family also explained in chapter 3. The interesting findings of the jumping spider include the high sensitivity with the defocusing system, which allows them to hunt at a further distance much more than their own body length. In this depth range, the Zf value does not matter since a stereo approach is suggested. Therefore, during the training process, the Zf value grows larger due to the algorithm racing on the DFDD approach but failing.

In the (30mm - 70mm) and (80mm - 120mm) depth ranges, a promoted camera design enlarges the aperture size and hence adds some penalty on the stereo baseline. Recall Figure 5.4 for the depth range of (20mm, 500mm), it appears that both stereo and DFDD methods get the minimum in some of the depth values. Similarly in this case, since in both depth ranges, either stereo or DFDD approaches have a lower error in some depth values and thus the final model appeared as the second and third blue images in Figure 7.1. Both cases have a stereo baseline and a not minimum right aperture. Comparably, the depth range (30mm - 70mm) is nearer, and a lower error suggested a comparably larger stereo baseline. The depth range (80mm - 120mm) is relatively further and the aperture size is larger.

In the (100mm - 200mm) and (300mm - 500mm) depth ranges, the DFDD methods perform much better than the stereo approach. Therefore, the error grows as the stereo baseline gets larger and the right aperture shrink. In Figure the last two blue images in Figure 7.1, the plot shows a lens arrangement preference on a big aperture with a baseline of the right aperture's radius. The last figure in 7.1 also shows a wide open aperture is preferred. This well illustrates the defocusing system of jumping spiders who hunt in a far field much larger than their own body size.

7.3 Relating to Mobile Photography

Today, mainstream phones implemented multiple camera systems with different aperture sizes. For instance, iPhone 14 implemented the dual-camera system with a 12 MP $f/1.5$ camera and

a 12 MP $f/2.4$ aperture [30]. Also, the Google Pixel 7 has a 50MP $f/1.85$ as a main camera, 12MP $f/2.2$ as an ultrawide camera, and a 48MP $f/3.5$ camera as its rear camera system [31]. Also, a lot of phone brands are implementing more than two cameras, such as OPPO Find x5 and Samsung Galaxy S23 [32, 33]. Those cameras are implemented with different aperture sizes to reach a different Depth of Field (DoF) of the target scene. However, which aperture sizes of cameras should we include? How to position them together so that they can work for the best depth perception together?

We can correlate wolf and crab spiders' stereopsis and jumping spiders' defocusing visual system to mobile photography! As explained, current mobile phones also mount cameras with different aperture sizes in order to use the stereo approach to obtain depth information and achieve different focusing depth ranges for shooting various DoFs. With multiple cameras implemented in mobile phones, we can take advantage of this model and come up with the best camera arrangement that is able to work to jointly achieve accurate depth estimation in a wider depth range. When a target is within the approximated depth range, the camera system is able to connect an appropriate pair of cameras to reach the best depth perception result.

CHAPTER 8

CONCLUSION AND FUTURE WORK

This model is designed to detect how the difference between wolf spiders and crab spiders using stereopsis and focusing system and jumping spiders utilizing stereopsis and defocusing technique. We explain this difference in their visual system by analyzing how their eye behave when observing targets at different depth ranges. Knowing more about the unique visual specialties of the three families of spiders, the vision system of jumping spiders can also be used as a novel inspiration for mobile camera designs. A wider depth range can be reached through a good arrangement of lenses when combining stereo methods which have higher accuracy in the near field, and DFDD which provides higher sensitivity in far vision.

Communication between those lenses through both stereo and DFDD approaches, a more complex camera model, when more lenses are implemented, can select the best positions and aperture sizes of the camera(s) to offer the best in-focus polished images. Moreover, a more accurate mobile depth estimation can also enhance its ability to plot depth maps and help to create a more friendly platform for various mobile AR applications [34, 35, 36].

This model is also open-sourced at Google Colab [37] at Spider_Eye_Behavior.ipynb for others to use on hunting spider vision research or camera system design(re-train on different loss function or error metrics for the model is needed).

8.1 Equation Optimization

A more accurate depth estimation function is presented as

$$\varepsilon \leq \frac{Z|Z-Z_f|}{Z_f}\sqrt{4\frac{\varepsilon_\omega^2}{\omega^2}+\frac{\varepsilon_B^2}{B^2}+\frac{\varepsilon_{Bt}^2}{B_t^2}}$$

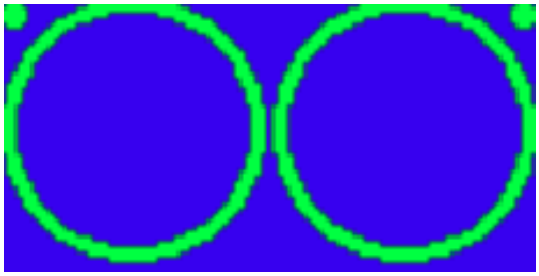

Figure 8.1: Expected Final Model Result

in Alexander's work [2]. However, since we do not have specific camera information to test on this algorithm, unknown parameters such as ε_{Bt}, ω, ε_B, B, ε_{Bt}, and B_t are not ready to be implemented at this stage. If more specific camera information is added, this model can produce more accurate DFDD error metrics.

8.2 Algorithm Optimization

The current algorithm supports the lens arrangement in one scale in which both baseline and right aperture are only arranged constraining their length as shown in Chapter 6. However, I also wish to constrain the width information of the lenses. This added feature can make the lens arrangement in a 2D scale as figure 8.1 while more cameras are added into the training. Thus, this approach of combining stereo and DFDD can be practical for industrial camera designs. Meanwhile, this implementation allows designs of more than two lenses and thus can achieve more complex camera models.

REFERENCES

[1] E. M. Ullrich-Lüter, S. Dupont, E. Arboleda, H. Hausen, and M. I. Arnone, "Unique system of photoreceptors in sea urchin tube feet," *Proceedings of the National Academy of Sciences*, vol. 108, no. 20, pp. 8367–8372, 2011.

[2] E. Alexander, "A theory of depth from differential defocus," Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2021-09-28, Ph.D. dissertation, 2019, p. 117, ISBN: 9798684608841.

[3] R. T. Held, E. A. Cooper, and M. S. Banks, "Blur and disparity are complementary cues to depth," *Current biology*, vol. 22, no. 5, pp. 426–431, 2012.

[4] S. Riddle, "How bees see and why it matters," *Bee Culture: The Magazine of American Bee Keeping*, 2016.

[5] Y. M. Song *et al.*, "Digital cameras with designs inspired by the arthropod eye," *Nature*, vol. 497, no. 7447, pp. 95–99, 2013.

[6] L. S. Roth, L. Lundström, A. Kelber, R. H. Kröger, and P. Unsbo, "The pupils and optical systems of gecko eyes," *Journal of Vision*, vol. 9, no. 3, pp. 27–27, 2009.

[7] A. Toet and M. A. Hogervorst, "Portable real-time color night vision," in *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2008*, SPIE, vol. 6974, 2008, pp. 13–24.

[8] K. Schechner, "Depth from defocus vs. stereo: How different really are they?" *International Journal of Computer Vision*, vol. 39, pp. 141–162, 2000.

[9] D. P. Harland, R. R. Jackson, and A. M. Macnab, "Distances at which jumping spiders (araneae: Salticidae) distinguish between prey and conspecific rivals," *Journal of Zoology*, vol. 247, no. 3, pp. 357–364, 1999.

[10] S. Aguilar-Arguello, D. Gerhard, and X. J. Nelson, "Distance assessment of detours by jumping spiders," *Current zoology*, vol. 66, no. 3, pp. 263–273, 2020.

[11] D. Korman, "Depth perception in hunting spiders (araneae:salticidae, lycosidae, thomisidae)," Ph.D. dissertation, 2019, p. 21.

[12] M. F. Land, "Structure of the Retinae of the Principal Eyes of Jumping Spiders (Salticidae: Dendryphantinae) in Relation to Visual Optics," *Journal of Experimental Biology*, vol. 51,

no. 2, pp. 443–470, Nov. 1969. eprint: https://journals.biologists.com/jeb/article-pdf/51/2/443/2376624/jexbio_51_2_443.pdf.

[13] T. Nagata *et al.*, "Depth perception from image defocus in a jumping spider," *Science*, vol. 335, no. 6067, pp. 469–471, 2012.

[14] M. Hallett, "John benjamin dancer 1812–1887: A perspective," *History of Photography*, vol. 10, no. 3, pp. 237–255, 1986.

[15] *Capturing photos with depth*.

[16] *Udepth: Real-time 3d depth sensing on the pixel 4*.

[17] T. Khakhulin, D. Korzhenkov, P. Solovev, G. Sterkin, A.-T. Ardelean, and V. Lempitsky, "Stereo magnification with multi-layer images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8687–8696.

[18] M. B. Mariappan, X. Guo, and B. Prabhakaran, "Picolife: A computer vision-based gesture recognition and 3d gaming system for android mobile devices," in *2011 IEEE International Symposium on Multimedia*, IEEE, 2011, pp. 19–26.

[19] *Taraxl - usb stereoe camera for nvidia gpu*.

[20] G. Bianco, A. Gallo, F. Bruno, and M. Muzzupappa, "A comparative analysis between active and passive techniques for underwater 3d reconstruction of close-range objects," *Sensors*, vol. 13, no. 8, pp. 11 007–11 031, 2013.

[21] S. G. Narasimhan, S. K. Nayar, B. Sun, and S. J. Koppal, "Structured light in scattering media," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, IEEE, vol. 1, 2005, pp. 420–427.

[22] C. Chen and Y. F. Zheng, "Passive and active stereo vision for smooth surface detection of deformed plates," *IEEE Transactions on Industrial Electronics*, vol. 42, no. 3, pp. 300–306, 1995.

[23] J. Zbontar, Y. LeCun, *et al.*, "Stereo matching by training a convolutional neural network to compare image patches.," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2287–2318, 2016.

[24] J. Watson, O. M. Aodha, D. Turmukhambetov, G. J. Brostow, and M. Firman, "Learning stereo from single images," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, Springer, 2020, pp. 722–740.

[25] Z. Rao, Y. Dai, Z. Shen, and R. He, "Rethinking training strategy in stereo matching," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[26] S. Im, H.-G. Jeon, S. Lin, and I. S. Kweon, "Dpsnet: End-to-end deep plane sweep stereo," *arXiv preprint arXiv:1905.00538*, 2019.

[27] A. P. Pentland, "A new sense for depth of field," *IEEE transactions on pattern analysis and machine intelligence*, no. 4, pp. 523–531, 1987.

[28] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[29] *Tf.gradienttape nbsp;: nbsp; tensorflow v2.11.0*.

[30] *Iphone 14 and iphone 14 plus*.

[31] *Google pixel 7*.

[32] *Oppo find x5 pro*.

[33] *The best camera features on your galaxy s23*.

[34] J. Paavilainen, H. Korhonen, K. Alha, J. Stenros, E. Koskinen, and F. Mayra, "The pokémon go experience: A location-based augmented reality mobile game goes mainstream," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17, Denver, Colorado, USA: Association for Computing Machinery, 2017, 2493–2498, ISBN: 9781450346559.

[35] R. A. G. Aparicio, J. J. R. Aliaga, and D. G. Q. Velasco, "Mobile application for the recommendation of furniture and appliances through augmented reality to improve the user experience in the online shopping process," in *Proceedings of the 2022 3rd International Conference on Internet and E-Business*, ser. ICIEB '22, Madrid, Spain: Association for Computing Machinery, 2022, 1–6, ISBN: 9781450397322.

[36] J. Qian *et al.*, "Portal-ble: Intuitive free-hand manipulation in unbounded smartphone-based augmented reality," in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '19, New Orleans, LA, USA: Association for Computing Machinery, 2019, 133–145, ISBN: 9781450368162.

[37] E. Bisong and E. Bisong, "Google colaboratory," *Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners*, pp. 59–64, 2019.