

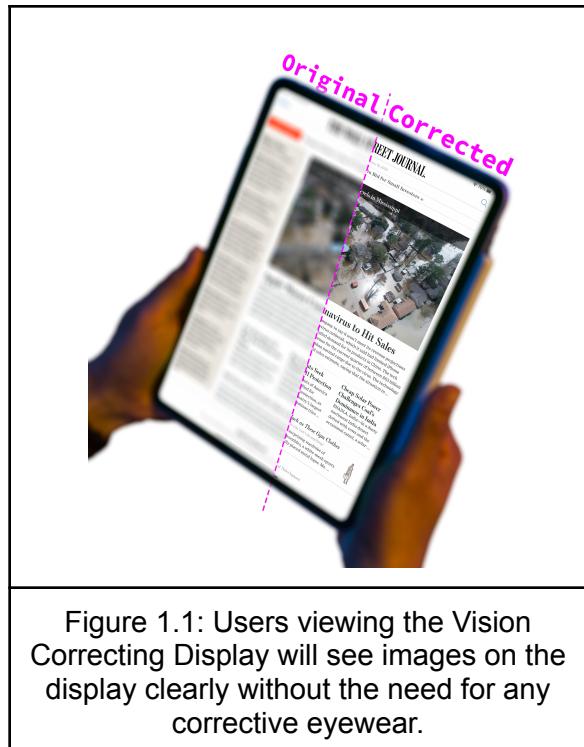
Microlens Array Subgroup Fall 2020 - Spring 2021 Project Report

Jimmy Xu, Angela Chen, Eric Ying

1. Motivation & goal

Vision problems such as near-sightedness and far-sightedness, are due to optical aberrations in the human eye. These conditions are prevalent, and the number of people who have these hardships is growing rapidly. Correcting optical aberrations in the human eye is traditionally done optically using eyeglasses, contact lenses, or refractive surgeries; these are sometimes inconvenient or not always available to everyone.

To tackle the issues of corrective lenses as mentioned above, the Vision Correcting Display (VCD) research group has utilized light field display or computational prefILTERING to help viewers perceive sharp images on the display without corrective eyewear. By introducing light field display technology and prefILTERING, the display is enabled to account for common vision defects and even higher-order aberrations. Besides, since the computation is performed in software, the display can be reconfigured for different users, which allows a single device to be tailored for multiple people.



Our primary goal this semester is to implement microlens-based algorithms. For the first iteration, we assume that the microlens is its focal length away from display, and the one-to-one assumption holds. For the second iteration, we break the assumption that the microlens is its focal length away from display. We then evaluate the results using various metrics, such as PSNR, HDR-VDP 2, and SSIM. We perform a series of experiments to try to get better results by using different sets of parameters.

2. Problem definition

In order to use vision-correcting displays, we must first compute a prefiltered version of any image to be displayed. This prefiltering process uses an abstracted eye model and ray-tracing to calculate the paths of light rays through the eye and determine where they land on the retina (if they do, in fact, reach it). We then compute the perturbed version of the original image that resolves to the correct image on the user's retina.

Previous works mainly focus on pinhole-based approaches. However, one drawback we noticed when using pinholes was the lack of brightness of the prefiltered image. Microlens light fields ideally should address that problem, since microlens arrays allow more light to pass through than pinhole masks.

Some results we get are (1) A working, generalized implementation of microlens-based VCD algorithm, (2) Comparison between the output images generated by pinhole-based and microlens-based VCD using a variety of metrics, and (3) Minimized distance between screen and microlens array (aka thickness of microlens array) while maintaining image quality and resolution.

3. Background

Diopter, myopia, and presbyopia

Diopter is the refractive power of the eye (its ability to bend light rays passing through), and is measured as the reciprocal of the eye's focal length. Typical relaxed human eyes have 60 diopters. As diopters increase, the focus length decreases, and light rays passing through the lens will converge closer to the lens. As diopters decrease, the focus length increases, and light rays converge further from the lens. A normal human eye adjusts its diopter dynamically depending on the distance of the object focused upon.

In myopia, the relaxed human eye has an abnormally high diopter, causing light rays to converge before the retina and creating circular blur. In presbyopia, the human eye loses the ability to accommodate when focusing on close objects, thus losing diopters. For the purposes

of simulating these aberrations, a single plane (or point) is used to separate the objects in focus versus the objects out of focus. In the case of myopia, there exists a far plane such that objects further away are out of focus. In the case of presbyopia, there exists a near plane such that objects closer are out of focus.

Nearsightedness	Far Point (mm)	Farsightedness	Near Point (mm)
-2D (Myopia)	500	+1D (Presbyopia)	330
-3D (Myopia)	330	+2D (Presbyopia)	500
-5D (Myopia)	200	+3D (Presbyopia)	1000

Figure 3.1: Far point and near point specifications for each degree of myopia and presbyopia tested during the experiments.

Light field display, microlens array, pinhole mask

In a conventional display, each pixel emits light in all directions, and we can only change its color. In contrast, a light field display can control both the color and the direction of the light emitting from each pixel. The additional direction information is vital to our implementation of vision correcting display.

Previous research groups used mostly pinhole-based light field displays. A pinhole mask is a sheet with arrays of openings. It covers a conventional screen such that the pixels below each pinhole can only go through the pinhole in a certain direction, creating a light field. A microlens array is a plane made up of arrays of tiny lenslets. The pixels below each lenslet will be bent in a certain direction, creating a light field.

The microlens array allows more light to pass through but is significantly more expensive than pinhole mask. Figure 3.2 shows the difference between the two types of display.

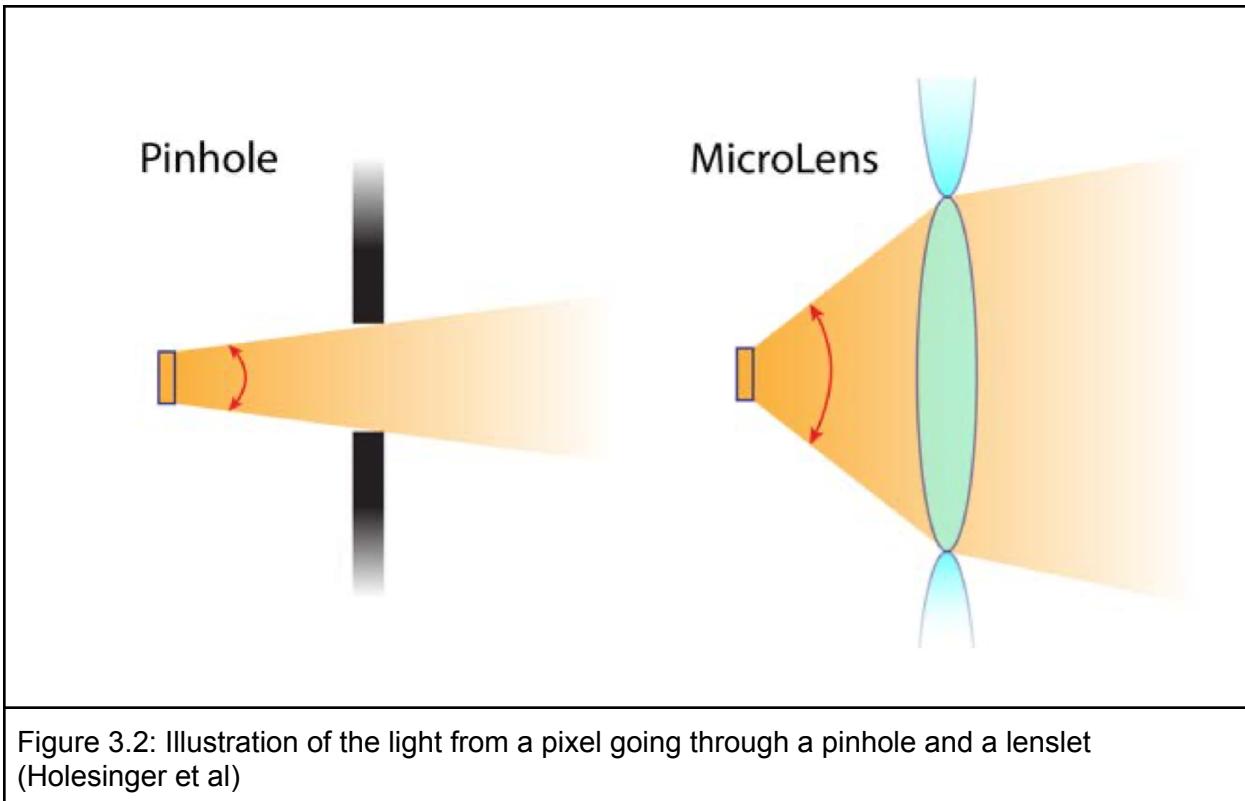


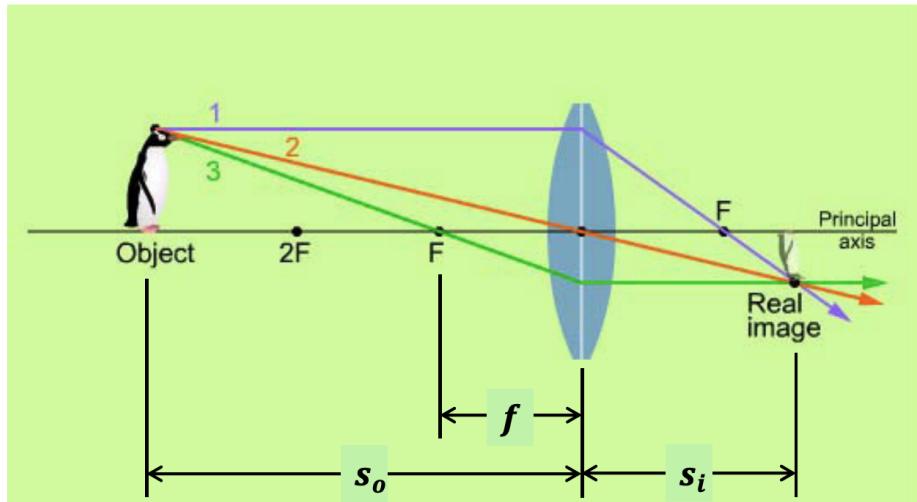
Figure 3.2: Illustration of the light from a pixel going through a pinhole and a lenslet (Holesinger et al)

One-to-one assumption

We assume there exists only one pinhole or microlens that the light from each pixel will go through and reach the eye. This assumption holds true in general.

Thin lens

A thin lens is a lens in which the thickness of the lens is much smaller than its diameter. We consider the thickness negligible when performing ray tracing, which significantly simplifies computation without causing too much error. Figure 3.3 shows some physical facts of thin lenses.



1. A ray parallel to the principal axis is refracted through the focal point.
2. A ray through the center of the lens is not refracted.
3. A ray through the focal point is refracted parallel to the principal axis.

Figure 3.3: Illustration of a thin lens
http://www.physics.purdue.edu/~jones105/phys21900_Fall2015/Phys21900_Lecture19.pdf

Forward vs. backward

The prefiltering process can be categorized as a “forward” approach or “backward” approach, based on whether the rays are traced from the display to the retina or vice versa, respectively.

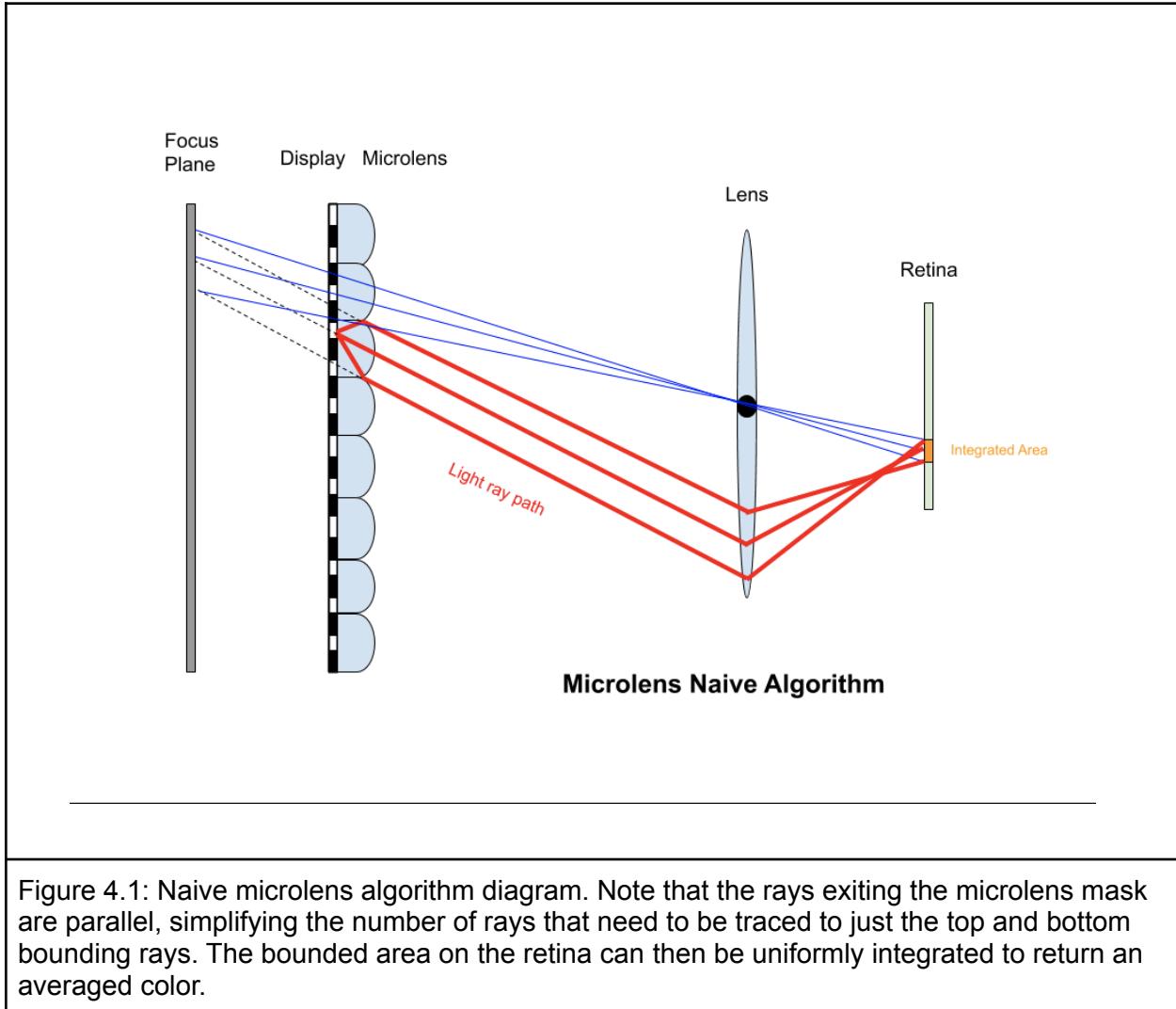
Depth of microlens array

The depth of a microlens array is defined as the distance between the microlens and the display.

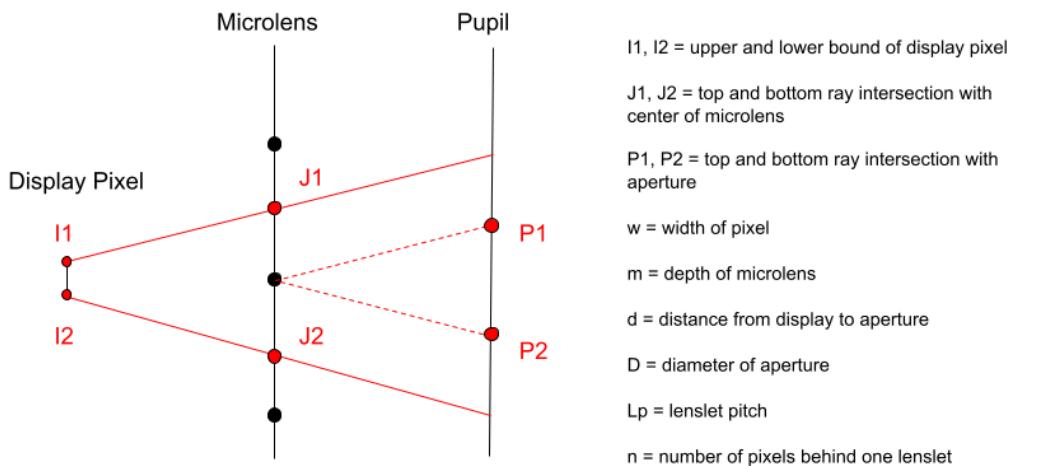
4. Proposed Approach

The first algorithm that we implemented was the **naive microlens forward algorithm**. Similar to the pinhole based forward algorithms (point to point, area to area), this algorithm traces rays from a display pixel to the retina and maps the colors of the retina pixels back to the display pixel. Our naive microlens algorithm assumes that the focal length of each individual lenslet in the microlens array is equal to the depth of the microlens array. This assumption allows us to

assume that each lenslet in the microlens array outputs a bundle of parallel light rays (derived from gaussian ray tracing). Thus, only the top and bottom bounding rays need to be traced in order to encompass the entire bundle of light rays, simplifying ray tracing calculations greatly and reducing the time to execute the algorithm.



In addition, we verified the one to one assumption for the microlens forward algorithm. The one to one assumption states that light rays from one display pixel can only pass through the pupil through one possible lenslet in the microlens array. If this assumption is broken, then we would have to consider all lenslets in the microlens mask when tracing rays from a particular display pixel, dramatically increasing the computation time of the naive algorithm by a factor n^2 (n is the dimension of the microlens mask). The one to one assumption allows us to circumvent this by selectively choosing one lenslet to project light rays through.



$|P_1 - P_2| > D \implies$ One to one assumption holds

$$P_1 = \frac{d}{m} * (J_1 - I_1) + I_1 - \frac{L_p}{2}$$

$$P_2 = \frac{d}{m} * (J_2 - I_2) + I_2 + \frac{L_p}{2}$$

$$|P_1 - P_2| = P_1 - P_2 = \frac{d}{m} * (J_1 - J_2 - (I_1 - I_2)) + (I_1 - I_2) - L_p$$

$$= \frac{d}{m} * (nw - w) + w - nw = \boxed{\left(\frac{d}{m} - 1 \right) * (n - 1) * w > D}$$

As long as this inequality holds, the one to one assumption holds as well

Figure 4.2: One to one assumption derivation for the naive microlens algorithm. As long as the red boxed inequality holds among the parameters of the user and display, no two rays exiting the same display pixel and entering different lenslets will pass through the pupil at the same time. We only need to consider one lenslet (the lenslet that intersects the ray starting from the display pixel to ending at the origin of the pupil) when tracing rays in the forward algorithm.

After implementing the prefiltering algorithm, we implemented the microlens based simulation. The purpose of the simulation software is to simulate how a user who suffers from a specific case of myopia or presbyopia would perceive a prefiltered image on the vision correcting display. Due to the lack of physical hardware, only a software based simulation can be used to test the quality of the vision correcting display.

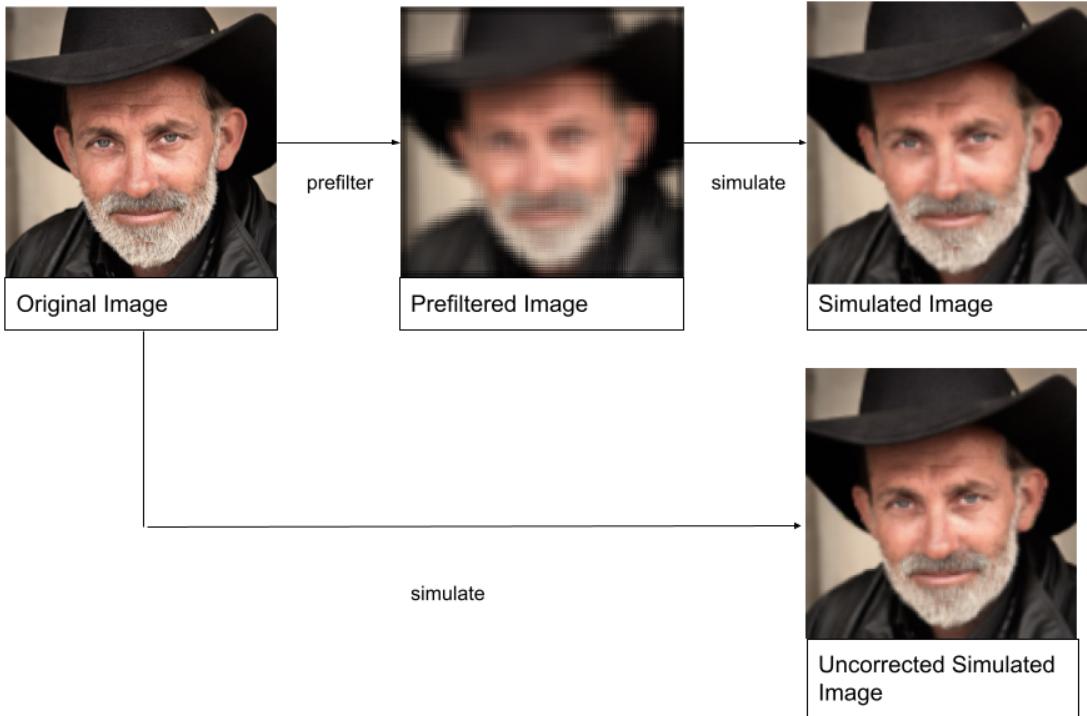


Figure 4.3: Combined prefiltering and simulation process. The prefiltering is responsible for distorting the image in some manner according to the user's prescriptions to display on the vision correcting display. The simulation is responsible for simulating how the user would perceive the image when viewing the display.

Previously, the simulation software targeted only pinhole based vision correcting displays. As a result, we had to extend the simulation software to target microlens based vision correcting displays. Like with the pinhole simulation, the microlens based simulation relies on a backwards ray tracing algorithm. For each pixel on the retina (which we can think of as the image that the user "sees"), multiple rays are traced backwards towards the prefiltered display. The colors of the display pixels hit are averaged together and mapped back to the retina. During this process, we make sure not to rely on any underlying assumptions on the user or display parameters, since the simulation must be generalized enough to apply to all possible user and display configurations.

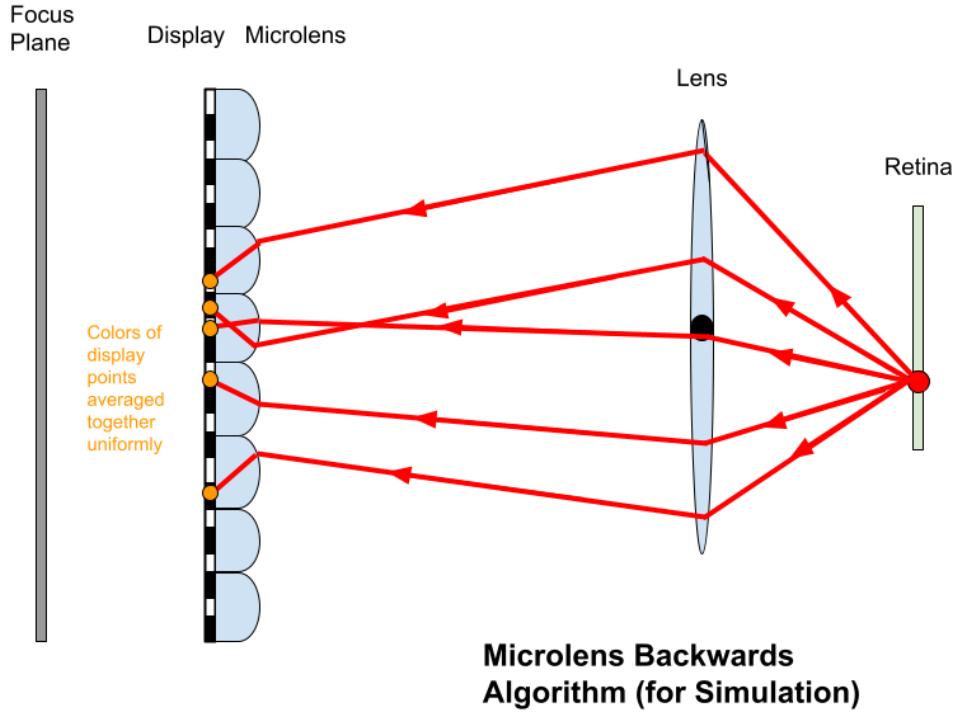


Figure 4.4: Microlens backwards simulation diagram. Notice that the rays are traced backwards from retina to display. Multiple intersection points on the aperture are sampled in order to retain as much information as possible for accurate simulation.

With the prefiltering and simulation in hand, our next objective was to verify the results of the algorithm against the same stock photos and image quality metrics used for the pinhole algorithms. This would be done whilst varying different hyperparameters. A secondary goal was to streamline the testing process to ensure it can be reused by future researchers.

The three metrics we used to evaluate image quality were PSNR, SSIM, and HDRVDP-2, roughly ordered by increasing complexity.

PSNR (or Peak Signal to Noise Ratio), is a simple metric that takes the ratio of the highest signal (in the case of most images, this would be 255 for the maximum RGB value) and the mean squared error noise (calculated by taking the square of the difference between a reference and test image). PSNR is typically measured logarithmically in decibels.

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

Figure 4.5: PSNR equation. MAX_I represents the peak signal and MSE is the mean squared error. Measured in dB. Image compression typically yields a PSNR of around 30-50 dB.

SSIM (Structural Similarity Index Measure) is a metric that takes into account luminance (mean intensity of the signal, which you can also think of as the brightness), contrast (variance of the signal normalized with luminance), and structural similarity (correlation of the signal normalized with luminance and contrast). These three comparisons are combined together to form the final similarity measure.

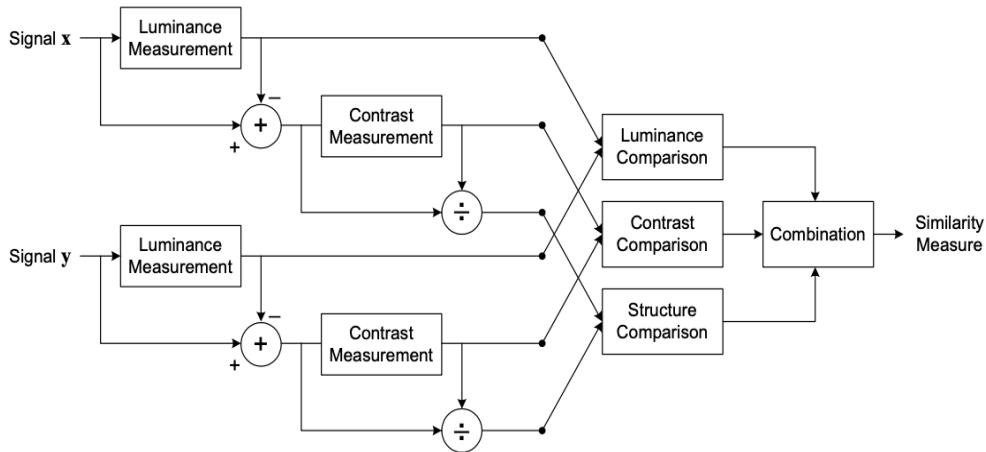


Fig. 3. Diagram of the structural similarity (SSIM) measurement system.

Figure 4.6: SSIM block diagram. The three comparisons are combined together to produce a final similarity measure.

HDRVDP-2 (High Dynamic Range Visual Difference Predictor) is the most complicated metric of the three, and it attempts to model the human visual system as closely as possible. HDRVDP-2 is composed of two metrics, the visibility metric, which measures how likely a human will notice the differences between two images, and the quality metric, which attempts to subjectively gauge how the visual differences between two images will affect the final visual quality perceived by humans. For the purposes of Vision Correcting Displays, we only concern ourselves with the quality metric, since we are interested in how well a user with a specific case of nearsightedness or farsightedness is able to perceive an image on the Vision Correcting Display.

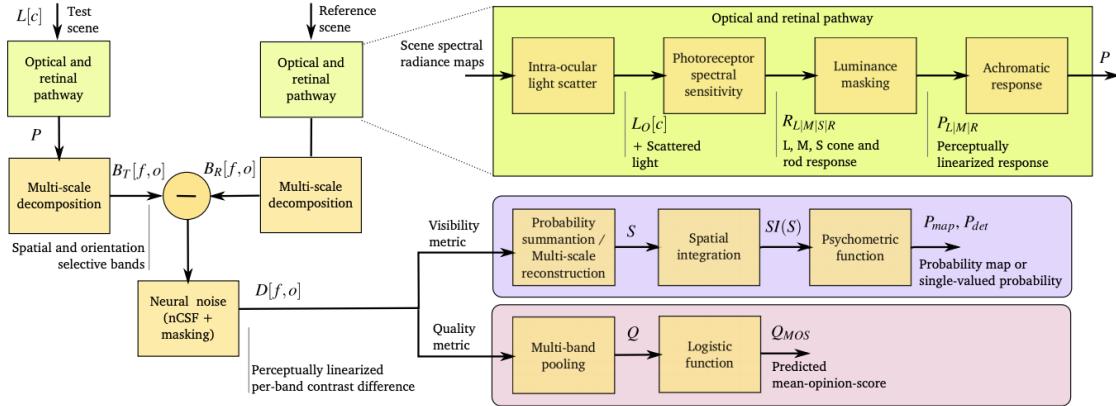


Figure 2: The block-diagram of the two visual metrics for visibility (discrimination) and quality (mean-opinion-score) predictions and the underlying visual model. The diagram also summarizes the symbols used throughout the paper.

Figure 4.7: HDRVDP-2 block diagram. Each stage in the pipeline attempts to account for some factor that influences the visual perception of an image.

Once the naive algorithm had been tested sufficiently, the final goal of this subgroup was to improve it. The proposed improvement was to break the assumption that the depth of the microlens array is equal to the focal length of lenslets. There are many theoretical benefits to doing this. Firstly, breaking this assumption allows us to reduce the thickness of the display for practical purposes. Secondly, changing the depth to be different from the lenslet focal length allows us to tune the parameters such that bundles of light exiting the display converge on the retina rather than scatter in an area, potentially mitigating focus blur.

$$\frac{1}{m} = \frac{1}{f_L} + \frac{1}{j - t},$$

Figure 4.8: The above equation derives the optimal depth of the microlens mask to minimize focus ambiguity from the thin lens formula. Here, m is the depth of the mask, f_L is the focus length of the lenslet, j is the desired projected bundle of light rays, and t is the distance between the microlens array and eye. The idea is we can tune m (the depth of the mask), assuming that everything else is constant, such that light rays hitting the user's retina are focused rather than scattered.

This motivated the creation of a new **generalized microlens forward algorithm**, which relies on a refraction function that maps input light rays hitting a lens to outgoing light rays exiting it. With this function, we can map light rays from the display plane to the microlens layer, refract towards the aperture/pupil layer, and then refract onto the retina. Since we can no longer assume that the bundle of light rays is parallel, we must discretely sample multiple light rays from a display pixel in order to retain as much information as possible. The final runtime of the algorithm is slower than the naive algorithm by a factor of k^2 , where k is the sampling rate on both the display and microlens layer (which can be adjusted).

Refraction calculation of thin lens

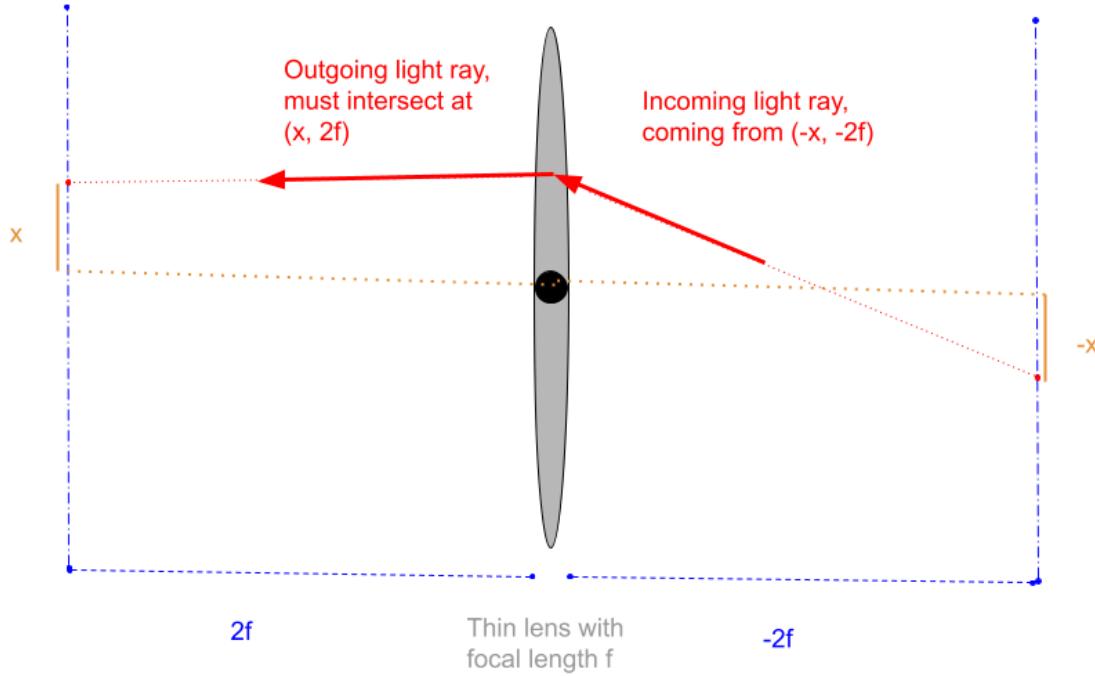
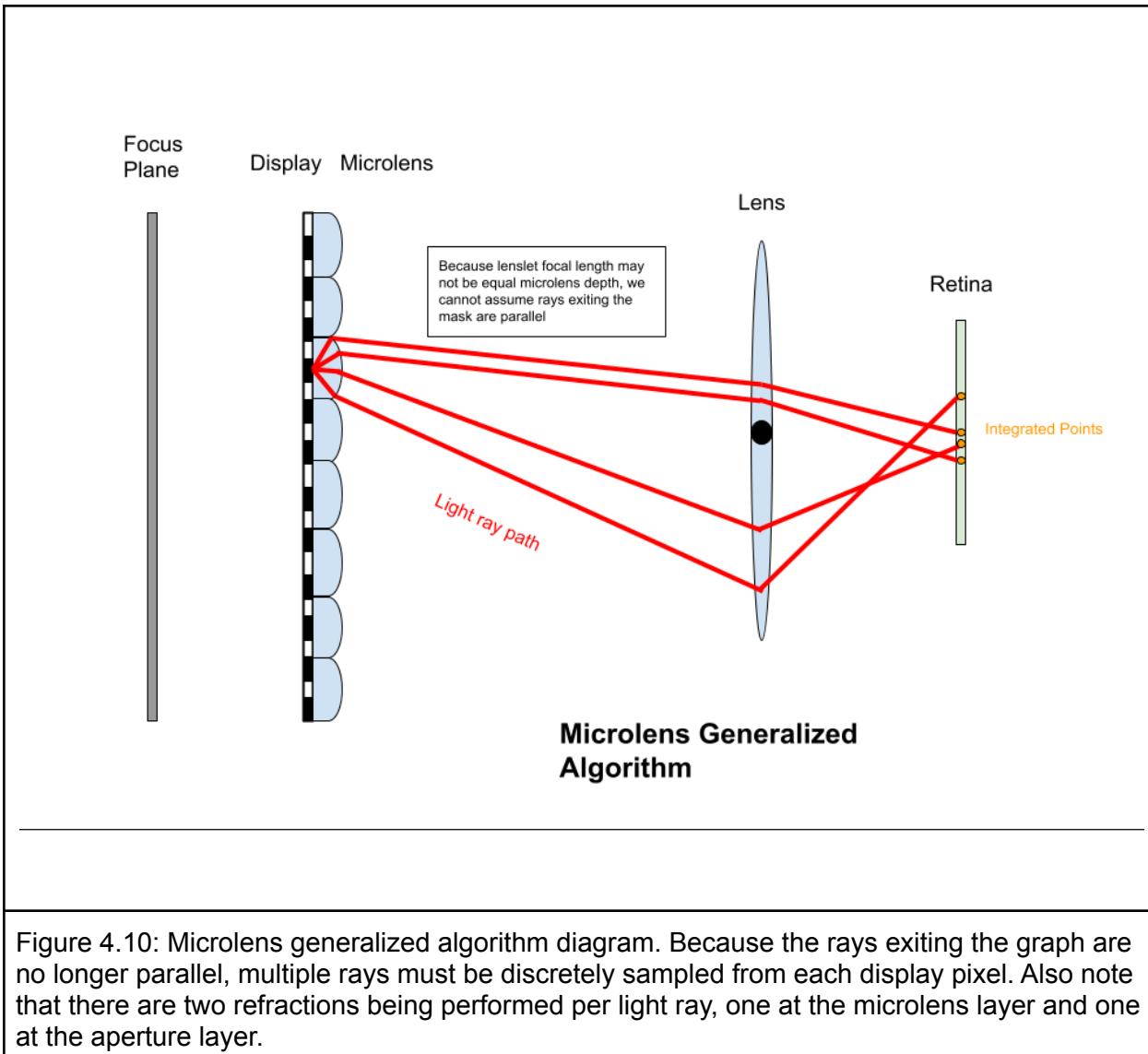


Figure 4.9: How to calculate the refraction of light rays across a thin lens. The incoming light ray is projected to $-2f$ away from the lens in the direction opposite of the incoming ray to yield a coordinate $-x$. The outgoing ray must then pass through $x 2f$ away on the opposite side of the lens by the thin lens equation ($\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}$).



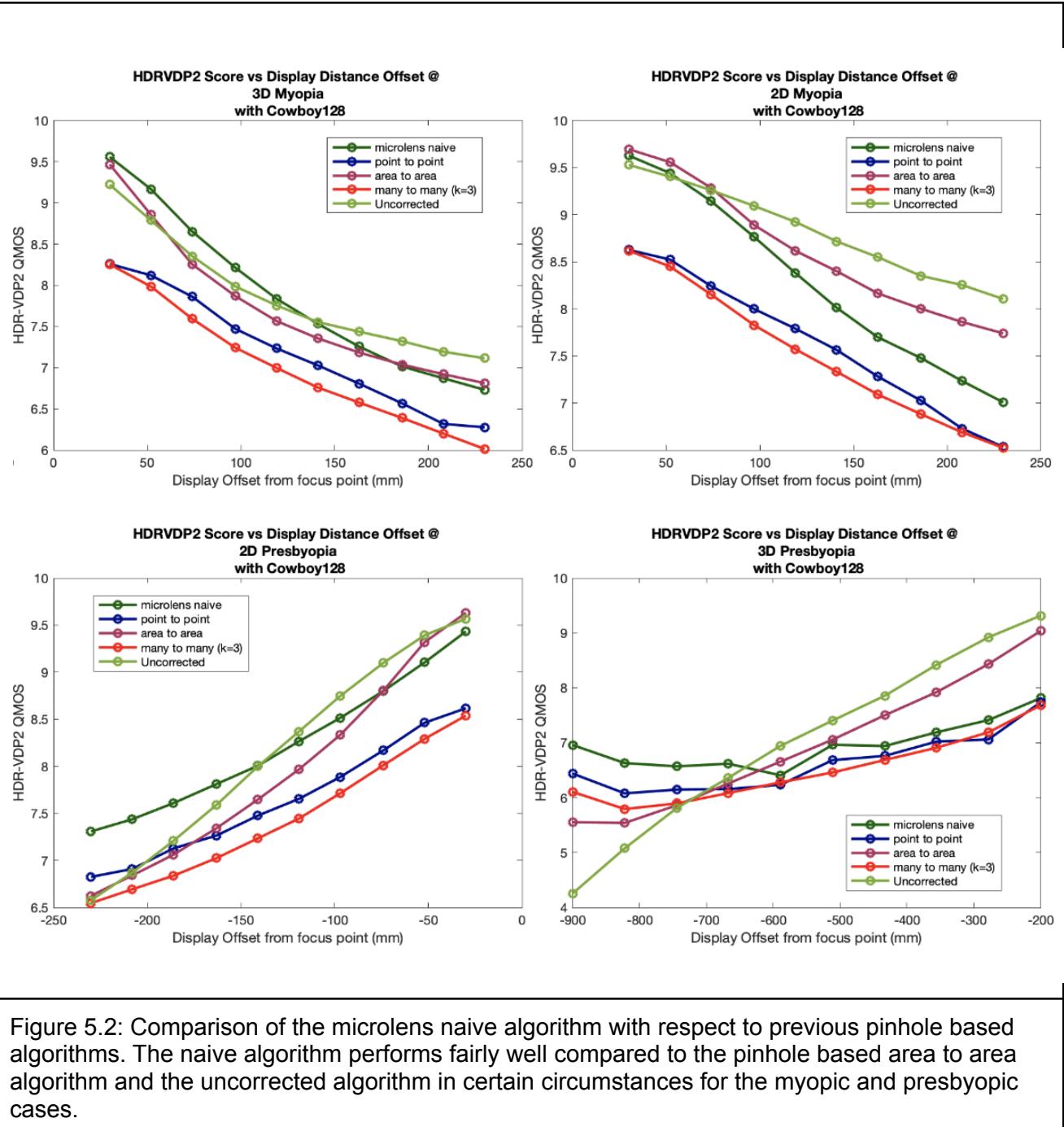
5. Results

Here will list some results gathered over the past semester. All following experiments use the below 128x128 cowboy reference image. This image is scaled up to 640x640 on the display during prefiltering and scaled back down to 128x128 on the retina during simulation.



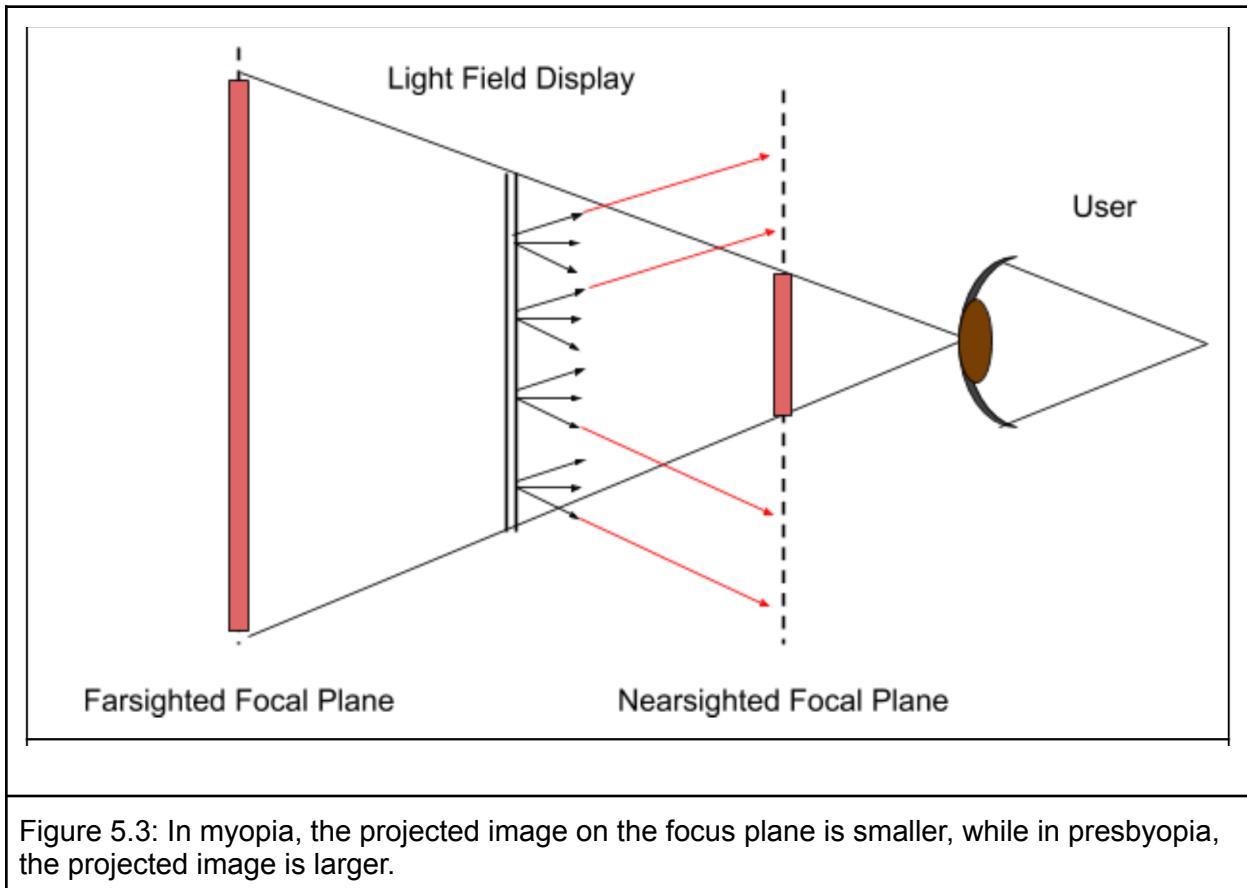
Figure 5.1: 128x128 cowboy reference image used for experiments. This image is scaled up during the prefiltering to 640x640.

We first compared the quality of the simulated images from the naive microlens algorithm with the simulated images from the pinhole based algorithms (which includes point to point, area to area, and many to many algorithms). The pinhole based forward optimization algorithm, which essentially solves the vision correcting problem posed as a least squares problem, was not included in the test suite due to its slow runtime. In addition, only the HDRVDP-2 results were included in this report, but the PSNR and SSIM results all followed the same pattern as the HDRVDP-2 results.



In the myopic (nearsighted) cases, the microlens naive algorithm performs fairly well with respect to the uncorrected results when the display is closer to the focus plane. In the presbyopic (farsightedness) cases, the microlens naive algorithm performs better when the display is further away from the focus point. One possible explanation regarding this phenomenon is that during prefiltering, the image at the display plane is essentially projected and rescaled onto the focus plane. In myopia, the rescaled image is smaller than the original

image. Thus, as the display moves further away from the focus plane, the rescaled image at the focus plane gets smaller, and less light rays contribute to the final prefiltering. In presbyopia, the rescaled image is larger. Thus, as the display moves further away from the focus plane, the rescaled image gets larger, and more light rays contribute to the prefiltering.



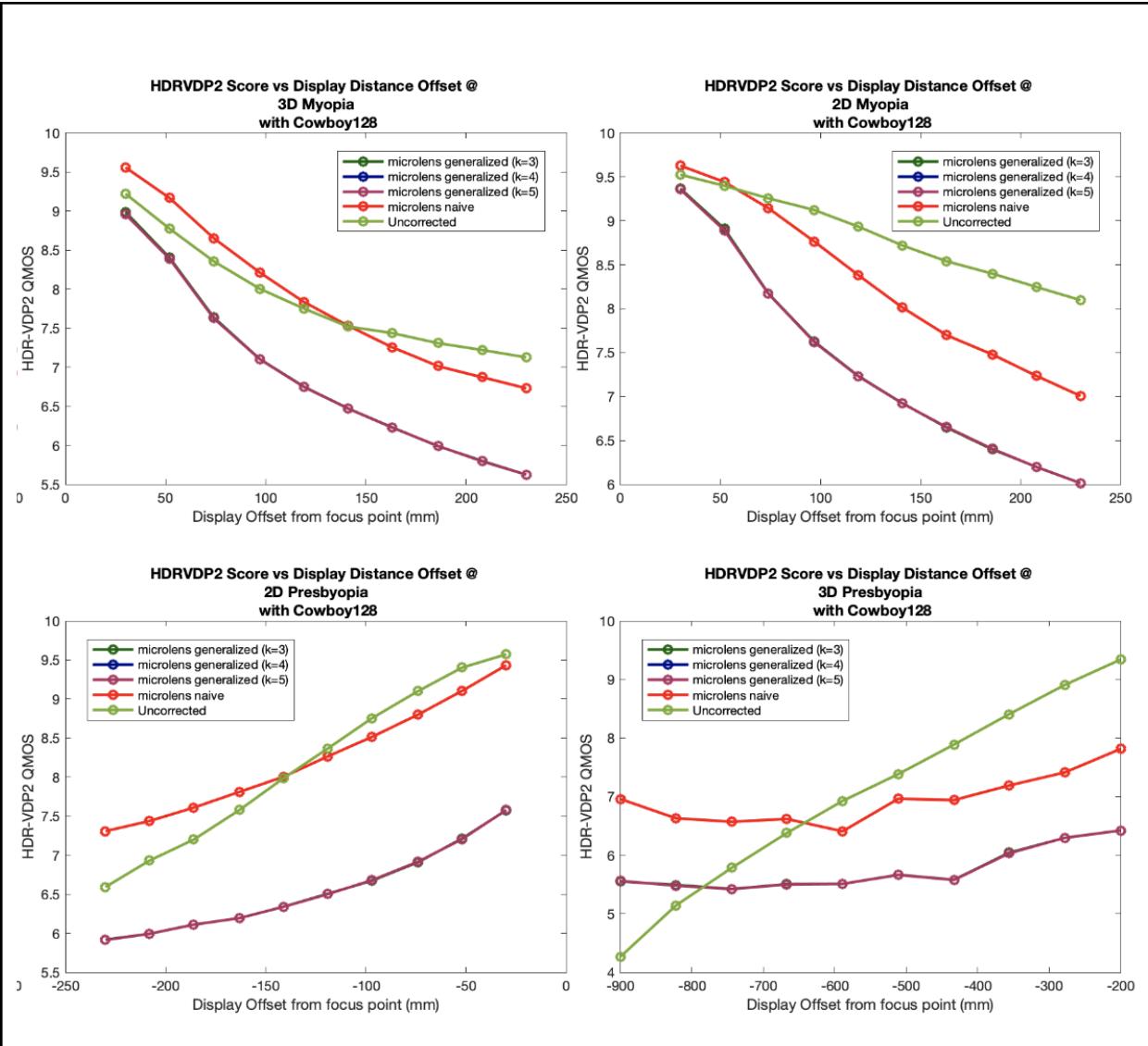


Figure 5.4: Comparison of the naive microlens vs generalized microlens algorithm. The generalized algorithm performs significantly worse than the naive algorithm. In addition, also note that sampling has almost negligible effect on image quality. The curves representing the generalized algorithm with sampling rate $k=3,4,5$ are stacked on top of one another.

On the other hand, the generalized algorithm performs consistently worse than the naive algorithm. One potential hypothesis that was explored is that the sampling rate is simply not high enough, and there is not enough information to contribute to the prefiltering process. However, higher sampling rates had a very negligible effect on the image quality. Another potential hypothesis was that the angular resolution is not high enough, but varying the angular resolution also had a very negligible effect on the final results. All this suggests that discrete sampling (like in pinhole many to many and microlens generalized algorithms) as a whole performs worse than continuous sampling (like in pinhole area to area and microlens naive algorithms). Further work needs to be done to verify this claim.

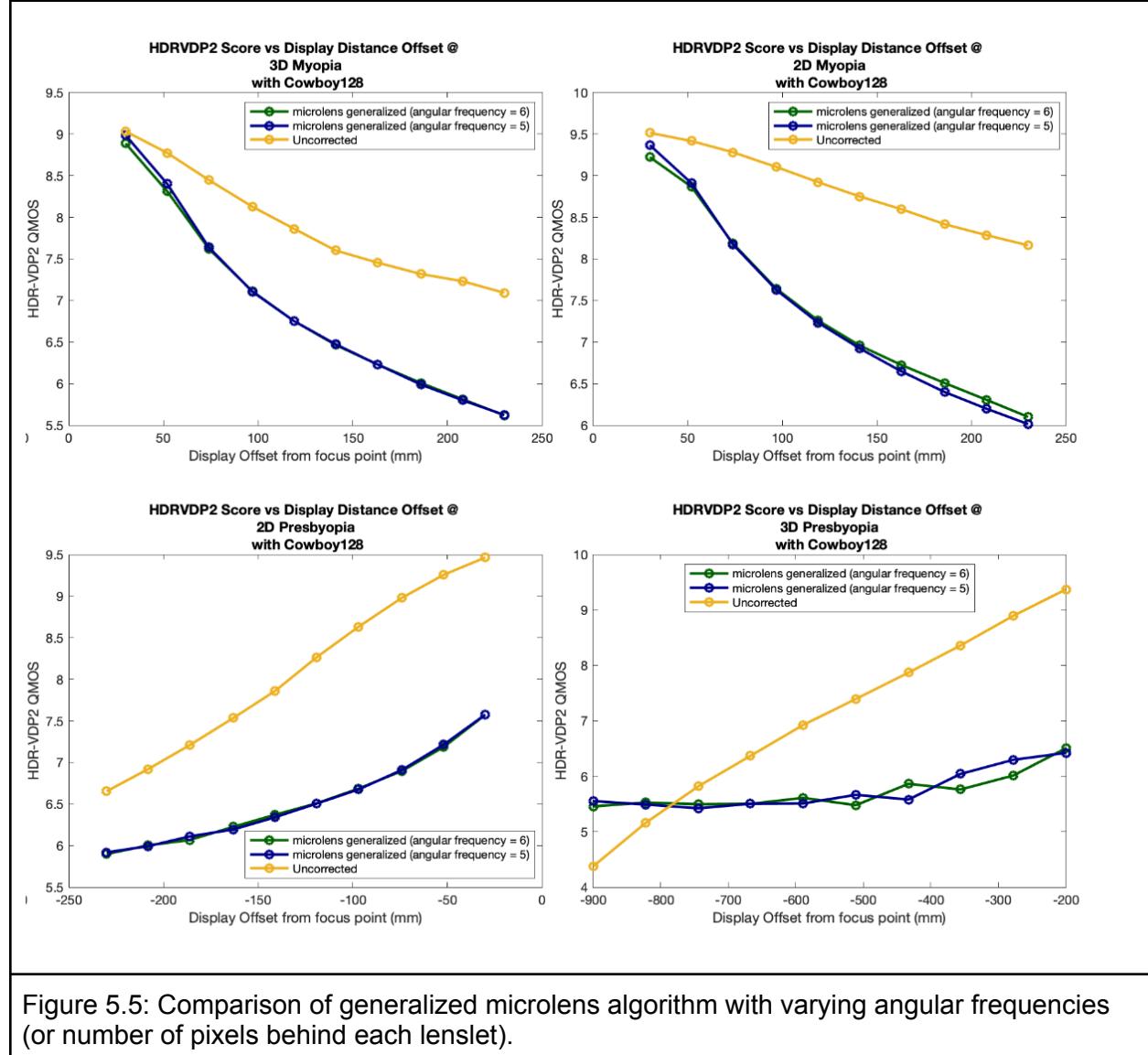


Figure 5.5: Comparison of generalized microlens algorithm with varying angular frequencies (or number of pixels behind each lenslet).

The final set of experiments performed was to investigate our earlier hypothesis regarding the optimal depth of the microlens mask with respect to the lenslet focal length, distance between

display and eye, and eye focus point. This was done by running the image quality metrics while fixing the far point, display distance, and lenslet focal length and varying the microlens depth. From the preliminary results, the quality of the image peaks at 6mm, the same distance as the lenslet focal length. While the current results do not support our hypothesis regarding the optimal microlens depth, they do reveal the robustness of the naive microlens algorithm (since the assumption that is made in the naive algorithm does not seem to adversely affect the final image quality at all).

Defocus	-2D Myopia
Far point	500 mm
Display distance	530 mm
Lenslet focal length	6 mm
Theoretical optimal microlens depth	7.5 mm

Figure 5.6: Experiment specification for testing the optimal depth of microlens mask. In the experiment, the depth of the microlens was varied from 5-10 mm.

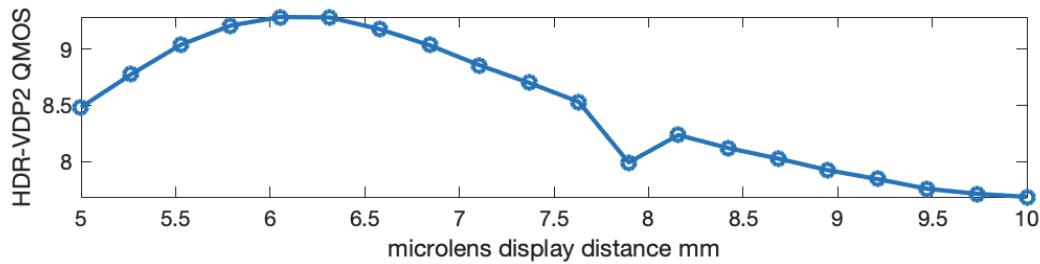


Figure 5.7: HDRVDP-2 quality of simulated image with varying microlens depth. The optimal depth to supposedly minimize focus blur is 7.5mm.

In conclusion, all experimental results seem to indicate that the naive microlens algorithm performs better than expected, whereas the generalized algorithm performs worse. Of course, further work needs to be done to verify the correctness of the generalized microlens algorithm and to investigate into the issue of focus blur.

6. Future Direction

Another aspect of the algorithm that the subgroup has been exploring is varying the indices of refraction. The previous prefiltering and simulation rely on a thin lens assumption for both the human eye and microlens array, which simplifies ray tracing formulas. When thick lenses are introduced, the index of refraction of the material used to create the microlens must be taken into account as well. Gullstrand's formula will be useful to look into and implement as it gives the focus length of a thick lens. Increasing the index of refraction will allow the thickness of the microlens layer to decrease while still providing the same refractive power. Incorporating the thick lens model and indices of refraction into the prefiltering and simulation model would allow for a more realistic and flexible Vision Correcting Display model. Future exploration should also be done into contemporary microlens array technology and the acceptable range of parameter values.

If there is to be greater exploration in differing depth and lenslet focal length, there would need to be further investigation into why discrete sampling (like in pinhole many to many and microlens generalized algorithms) performs worse than continuous sampling (like in pinhole area to area and microlens naive algorithms). More investigation may provide insights into why the microlens generalized algorithm performs much worse than the microlens naive algorithm and may lead to improvements in the microlens generalized algorithm that can help with fine-tuning the parameters.

Finally, the forward optimization algorithm can easily be extended to include microlens arrays, since we can apply the same projection conversion framework on microlens arrays as with pinhole masks. Incorporating microlens arrays may improve performance, because more light rays are projected from the display to the retina.

7. Related Work

Al-Kazily, T., Mao, M., Liu, E., Tsao, C., Hwang, W. C., & Ying E. (2020). *Algorithms for Displays that Correct Nearsighted and Farsighted Vision*. Electrical Engineering and Computer Sciences in the Graduate Division of the University of California, Berkeley

Bekerman, I., Gottlieb, P., & Vaiman, M. (2014). Variations in Eyeball Diameters of the Healthy Adults. *Journal of Ophthalmology*, 2014, 1-5. doi:10.1155/2014/503645

Holesinger, J., Barsky, B. A., & Kanté, B. (2020). Adapting Vision Correcting Displays to 3D (Issue UCB/EECS-2020-181) [Master's Thesis, EECS Department, University of California, Berkeley].
<http://www2.eecs.berkeley.edu/Pubs/TechRpts/2020/EECS-2020-181.html>

Mantiuk, R., Kim, K. J., Rempel, A. G., & Heidrich, W. (2011). Hdr-Vdp-2. *ACM SIGGRAPH 2011 Papers on - SIGGRAPH* 11. doi:10.1145/1964921.1964935

PAMPLONA, V. F., Oliveira, M. M., Aliaga, D. G., & Raskar, R. (2012). Tailored displays to compensate for visual aberrations. *ACM Transactions on Graphics*, 31(4), 1-12. doi:10.1145/2185520.2185577

Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4), 600-612. doi:10.1109/tip.2003.819861

Zhao, Y. (2020). "Vision Correcting Displays With Microlens Array," Electrical Engineering and Computer Sciences in the Graduate Division of the University of California, Berkeley.

Zhen, Y. (2019). "New algorithms for the vision correcting display," Electrical Engineering and Computer Sciences in the Graduate Division of the University of California, Berkeley.