Dual pixel auto-focusing

Team members: Itamar Terem

Contents:

Introduction:

Camera focus can be done manually by rotating the lens focus ring or automatically by the system. An autofocus (AF) optical system uses a sensor, a control system, and a motor to focus on an automatically or manually selected point or area. Many auto-focusing techniques was developed throughout the years. For example, active AF systems [1] measure distance to the object independently of the optical system, and subsequently adjust the optical system for correct focus. On the other hand, passive AF system [1] determined the correct focus by performing passive analysis of the image that is entering the optical system. One example of passive AF system is Phase detection (PD) [2,3]. Phase detection auto-focus is achieved by dividing the incoming light into pairs of images (figure 1) and finding the disparity between them [4]. Objects that are in focus will have minimal disparity and blur compared to object that are out of focus. One way to divide the incoming light and generate the two images is done by using dual pixel type sensor (figure 2). The dual pixel-type sensor contains sub-pixels [5,6], which separates an input light path into two paths to generate the images. Several methods for calculating the disparity from the two images have been proposed. Among them, cross correlation methods based [7], multiscale features extraction and hierarchical motion estimation [8], and lately deep learning based [9]. So far, most of the auto-focusing algorithms were trained and examined on 2D scenes. In this work we implemented the dual pixel auto-focusing algorithm on a simulated 3D scene in ISET3D.

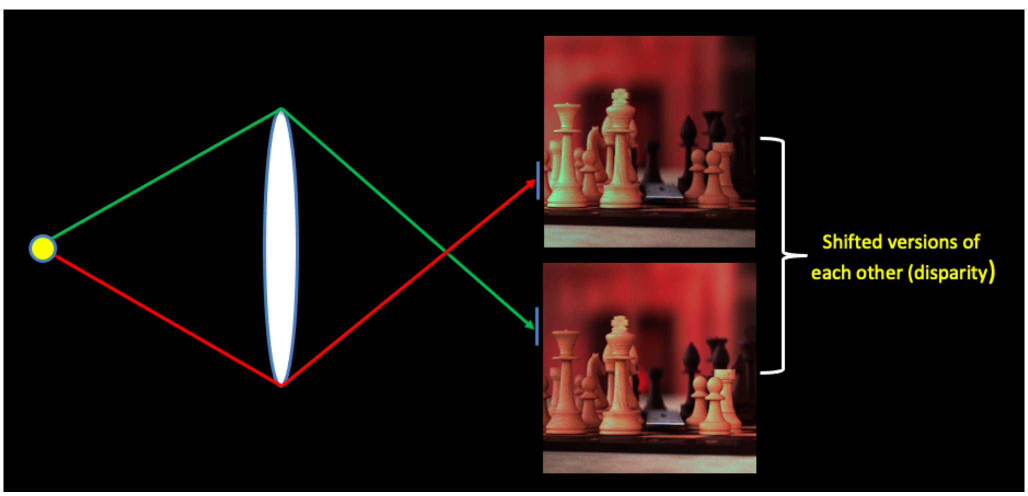

Figure 1: Phase detection (PD) methods are based on splitting the incoming light into pairs of images and estimating the disparity between them. Objects that are in focus, like the King piece, will exhibit low disparity in comparison to objects that are out of focus, like the Rook piece.

Background:

Dual Pixel Image Formation:

Dual Pixel are sensors where each pixel is split into two photodiodes sites (figure 2). Incident light is focus on a microlens, which splits the light into two pixels: left and right.

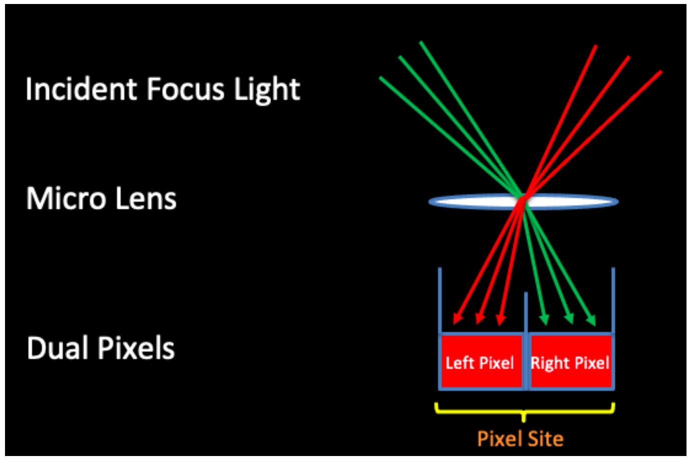

Figure 2: Incident light is focused into the microlens, which splits the light into two different photodiodes (left and right)

This split-pixel design enables the acquisition of two sub-aperture views [6], because each half of the DP accumulate light over one half of the aperture (figure 2). The two sub-views are kind of stereo pair, in which objects exhibit horizontal shift between the two sub-views (figure 3). The sum of the two views accounts for all the light going through the aperture and is equal to the ordinary full-pixel image that would be captured by a non dual-pixel sensor [5].

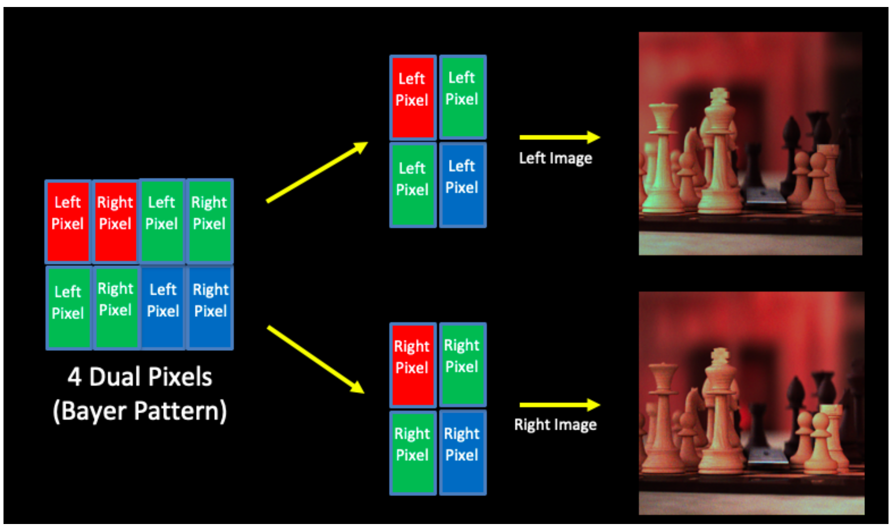

Figure 3: An example of the Bayer pattern dual pixel. The red, green, green, blue pattern consists of 8 photodiodes, where each per has it own microlens. This pattern enables to capture two images of the same scene from two sub-aperture views

Dual Pixel & Focus:

Figure 4 depict the effect of focusing on the forming two images. Objects in the scene that are out of the optical depth of field will be blurred and shifted (exhibit disparity) in the two sub-aperture views images. On the other hand, objects that are in focus (inside the optical depth of filed) will be sharp and have minimal disparity in the two form images (figure 5).

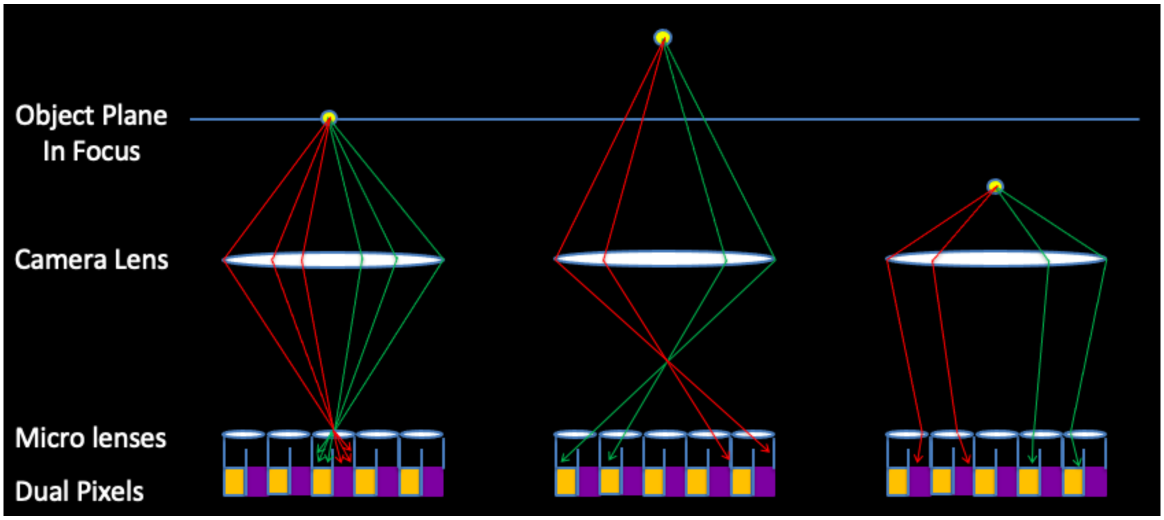

Figure 4: The affect of defocusing on the sub-views images. When a point source is placed at the object focal plane the camera lens focuses the light on the microlens, which splits it into to two different photodiodes (yellow and purple), with minimal shift and blur. On the other hand, when a point source is placed out of the object focal plane the light will no longer be focused on the microlens and the sub-aperture views images of the point source will be blurred and will exhibit disparity (shift) between them.

As a result, the disparity between the two views in a dual-pixel image is proportional to the defocusing of different objects in the full-pixel image. Phase Detection auto-focusing algorithms, which Dual pixel is one of them, take advantage of this phenomena in order to determine when different objects in the scene are in focus, by minimizing the disparity between the images.

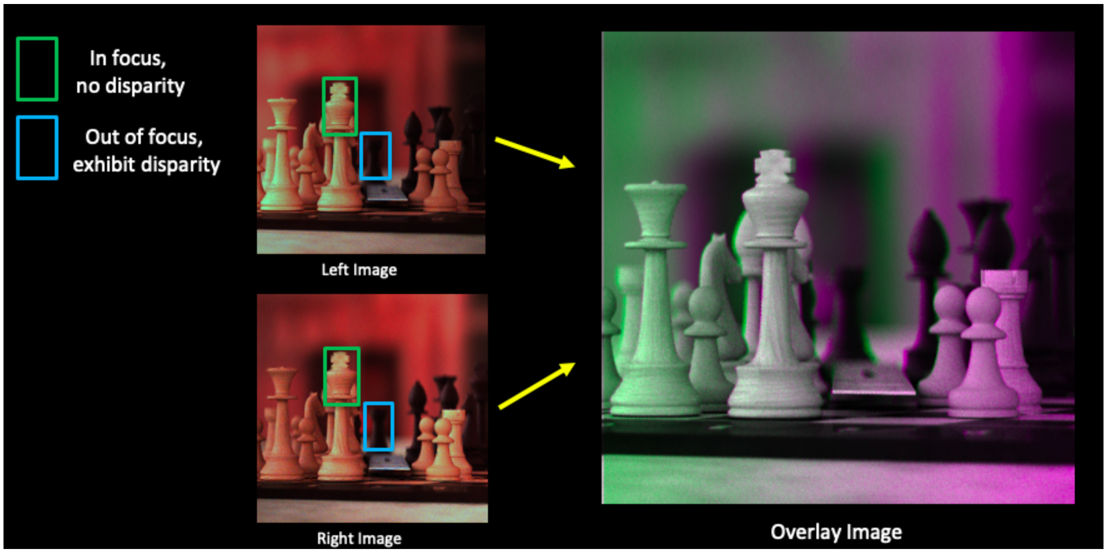

Figure 5: an example of the two sub-view images (left and right image), and their overlay image. The rook piece, which is out of focus, will exhibit disparity (as can be seen by the green and purple colors in the overlay image). On the contrary, the king piece which is in focus will exhibit no disparity (as can be seen by the absent of the green and purple colors in the overlay image)

Disparity Estimation:

As we saw in the section above, we need a way to estimate the disparity between the two sub-views images in order to determined when an object in the scene is in focus. One optional algorithm is Template Matching using Zero Mean Normalize Cross Correlation (ZNCC) [7]. In this algorithm a template image is shifted pixel by pixel in the horizontal and vertical directions across a reference image using the origins of the two images as a reference points (figure 6). For each shift the ZNCC function is evaluated (figure 7).

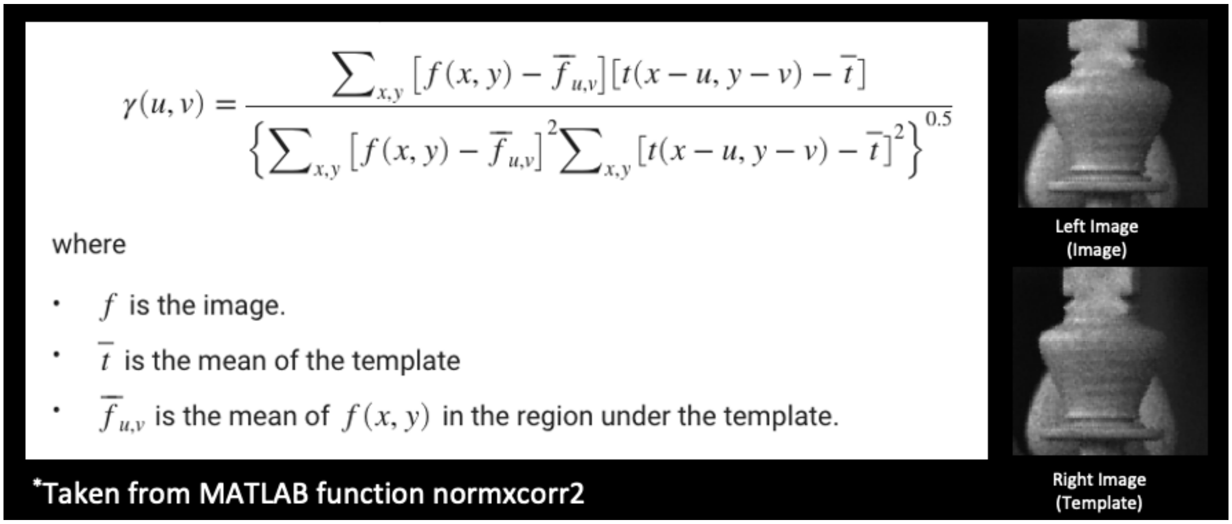

Figure 6: The Zero Mean Normalize Cross Correlation (ZNCC) function, the reference image (left image) and the template image (right image). The template image is shifted pixel by pixel in the horizontal and vertical directions across the reference image, and for each shift the ZNCC function id evaluated.

The maximum correlation occurs when the two images (template and reference image) are most similar to each other, which also corresponds to the location when the two are most align with each other. We can use this coordinated (of maximum correlation) in order to estimate the disparity. It is important to mentioned that this type of dual-pixel design result in one dimension (horizontal direction) shift, so only the X coordinate of the maximum correlation was used to estimate the disparity.

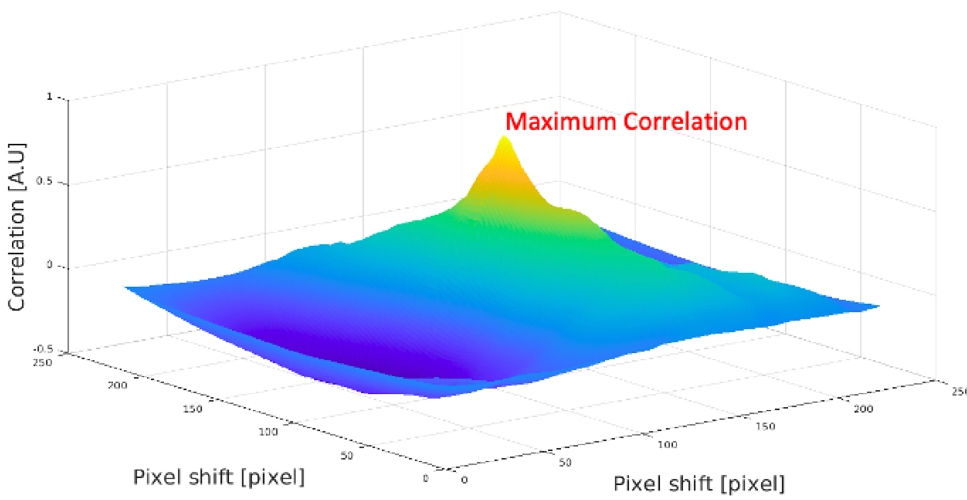

Figure 7: Zero Mean Normalize Cross Correlation (ZNCC) function. The disparity in the horizontal direction can be estimate from the X coordinate of the maximum correlation.

The ZNCC function is sensitive to low contrast and noise. This means that for low contrast template and reference images the algorithm will be more susceptible to errors in the disparity estimation. As a result, to improve the performance and robustness of the algorithm, the ZNCC function was evaluated over the RGB channel with maximum contrast. The contrast of each channel was calculated according to the following formula:

$$\frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$$

Script Outline:

As was stated in the introduction, the main goal of the project was the implementation of the dual pixel auto-focusing algorithm on a simulated 3D scene in ISET3D. The script at ISET3D contains the next following steps:

1. Uploading a PBRT recipe of a 3D scene (Slanted bar or Chess set).
2. Defining the optics to capture the scene (lens, aperture diameter, etc.)
3. Defining a dual pixel sensor and calculate the optical and digital image.
4. The user now can choose a region of interest (ROI) to focus on (figure 8)

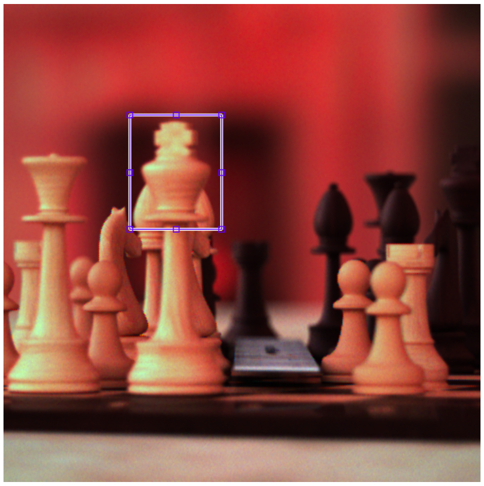

Figure 8: An example of the graphical user interface (GUI). The user is asked to choose a region of interest (ROI) to focus on

5. Looping over different focal planes to find the plane with minimal disparity (using the ZNCC) in order to determine when the object is in focus (figure 9)

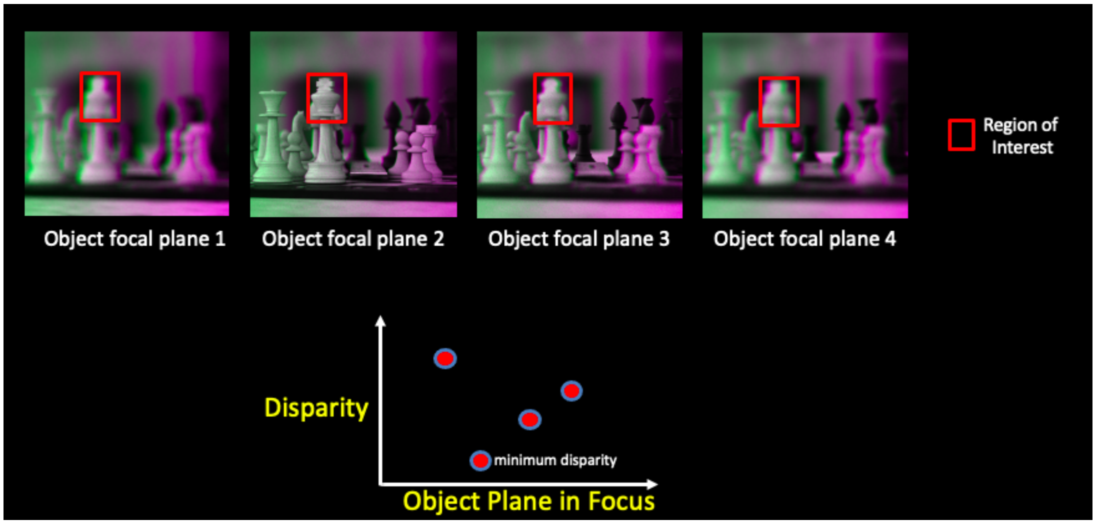

Figure 9: An example of the auto-focusing algorithm. The script generates digital images for different object focal planes, and for each of them calculate the ZNCC in order to estimate the disparity.

6. The algorithm then return the image with the chosen ROI in focus, by choosing the object focal plane with the minimal disparity (figure 10)

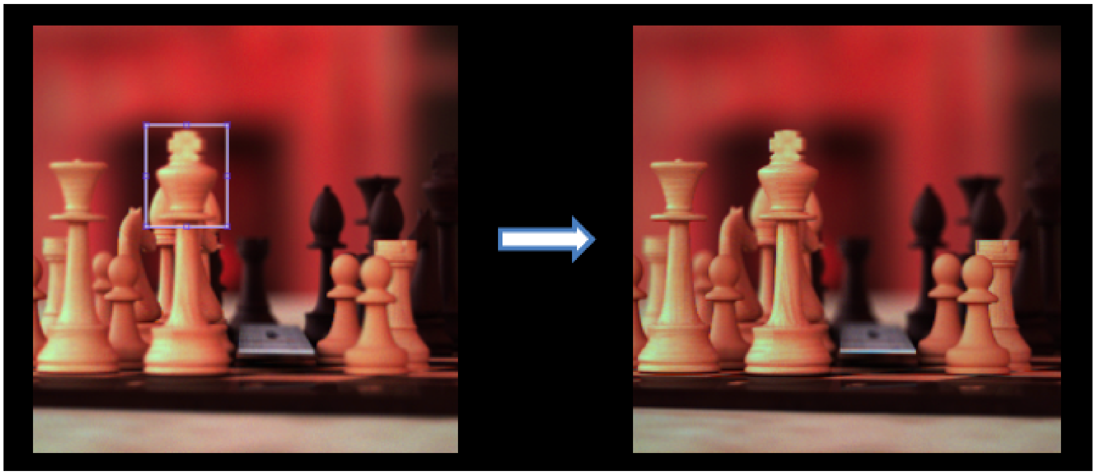

Figure 10: The script summary of the dual-pixel auto-focusing algorithm. The user chooses a region of interest (ROI) to focus on, and the algorithm automatically find where the object is in focus.

Improving Algorithm Robustness:

We noticed that sometimes the algorithms find the same minimum disparity value (zero shift) for adjacent focal plane (figure 15). This can happens because the ZNCC is sensitive to low contrast and noise. For example, the Rook background color is similar to the Rook color, which result in low contrast. To resolve this ambiguity we added another measure (image sharpness) in addition to the disparity estimation. The image sharpness was calculated for the summed image (left and right) using the following procedure. We take the cropped (over the ROI) summed image, and calculate the image magnitude gradient and image edges using the Sobel operator. Then, we multiply the two images to get the image magnitude gradients that correspond to edges, and calculate the mean (for non zero values) over the image. The larger the mean the sharper the image.

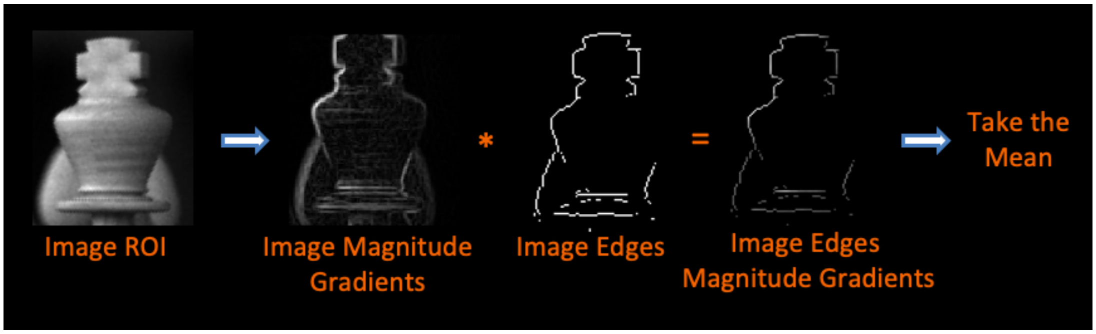

Figure 11: Image sharpness calculations. The image (cropped image) magnitude gradient and edges were calculated using the Sobel operator. The multiplication of the two result in an image magnitude gradients that correspond to the edges. Sharpness value is calculated by taking the mean (non zero values) over the result image.

To summarize, the algorithm starts by estimating the disparity of different object focal planes. Then, if ambiguity (same minimum disparity value for adjacent focal planes) occurs, the image sharpness is estimated, and the final object focal plane is chosen according to the maximum sharpness.

Results:

Slanted Bar:

A 3D scene with four slanted bar targets was created. Each target was put in different distance (2,3,4,5 meters) from the camera (to be more precise from the optical imaging film). The scene was capture using a double gauss 22 degree 50 mm lens, with an aperture diameter of 6 mm, and 50 rays per pixels. 256 by 256 microlens with a 2.8 um diameter, film diagonal of 1.0137 mm, and film resolution of 512 by 512 pixels (1.4 um pixel size) were chosen.

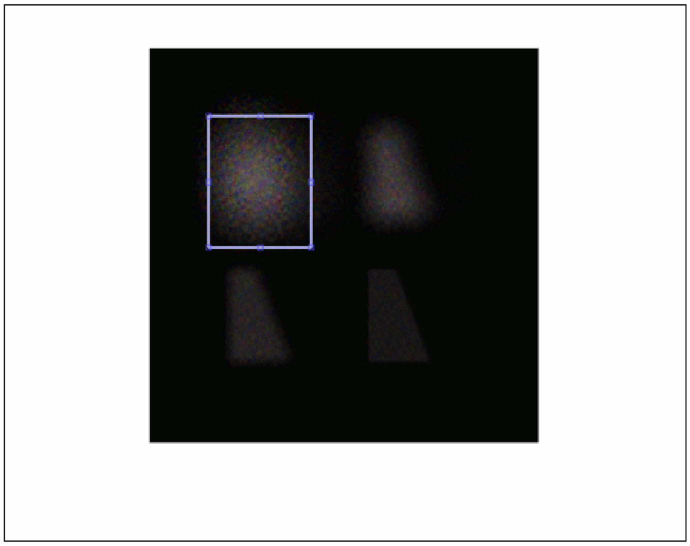

Figure 12: A GIF file demonstration of the dual pixel auto-focusing algorithm. The user first choose an ROI around the wanted target, and then the algorithm finds where the target is at focus. As can be seen, the algorithm succeeded to put each of the targets in focus.

The object focal plane was varying from 1 meter to 6 meters with 0.5 meter step (figure 12), while the algorithm goal was to place each target in focus automatically. Figure 13 depict the disparity result for each of the target as a function of focus distance. As can been seen, the minimum disparity (zero) for each target matched the correct distance (where the targets were placed). As was discussed in the section above (Dual Pixel & Focus) we expect the disparity to have larger values further away from the target focus. As can be seen, this is true for targets 1, 2 and 3, but not for target 4 (the disparity for 1 meter focus distance should be higher than 2 meters focus distance). Target 4 does not follow the correct trend, because the disparity estimation algorithm (ZNCC) is sensitive to low contrast. As can be seen in figure 12, at 1 meter focus distance the signal of the forth target is vert low, so the error in the disparity estimation is large. Nonetheless, this does not affect our final results.

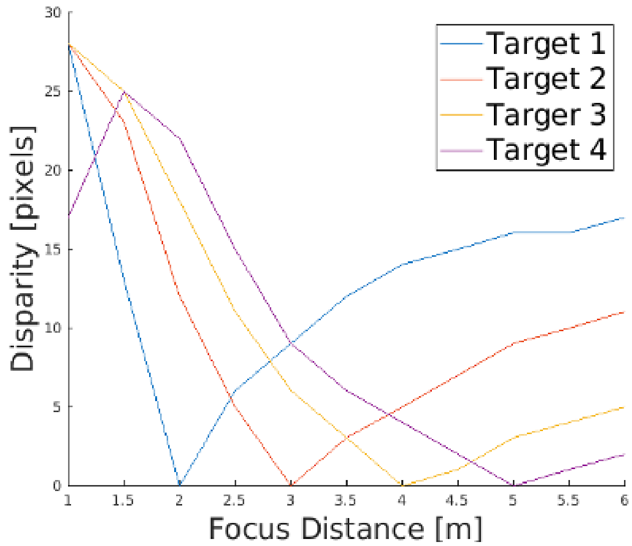

Figure 13: Disparity estimation as a function of focus distance for each of the four targets. As we can see the algorithm has succeeded to identify the correct object focal plane for all target.

Chess Set:

A 3D Chess Set 3D scene was captured using a double gauss 22 degree 50 mm lens, with an aperture diameter of 15 mm, and 200 rays per pixels. 512 by 512 microlens with a 14 um diameter, film diagonal of 10.1371 mm, and resolution of 1024 by 1024 pixels (7 um pixel size) were chosen. The imaging parameters were chosen to enable enough filed of view, with short depth of field in order to create a noticeable disparity effect.

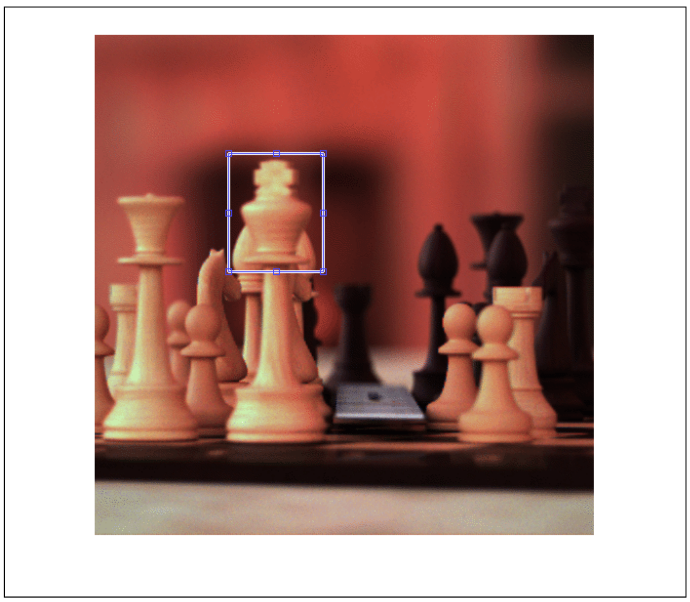

Figure 14: A GIF file demonstration of the dual0pixel auto-focusing algorithm on a real simulated 3D scene (Chess Set). An ROI around the wanted piece (Kind and Rook) was chosen, and the algorithm found where the targets were at focus.

The object focal plane was varying from 1 meter to 2.4 meters with 0.2 meter step (figure 14), while the algorithm goal was to place the Rook and Kind piece in focus automatically. Figure 15 depict the disparity result for each of the two pieces as a function of focus distance. As can be seen, a single minimum disparity (zero) was detected only for the King piece. This minimum corresponded to the correct object focus position. On the other hand, the algorithm estimate the same minimum disparity value for the Rook piece in two adjacent focus distances (1.6 & 1.8 meters). As was discussed in the section above (Improving Algorithm Robustness), this ambiguity was solved by estimating the cropped image (Rook piece) sharpness and choosing the one with the largest sharpness. One should also notice that the same disparity trend detected in the forth slanted bar plot is also exhibit in the Rook piece. The disparity at 1 meter focus distance should be higher than the one in 1.2 meter focus distance. As before, this error occurs due to the high sensitivity of the ZNCC function to low contrast and significant blurring.

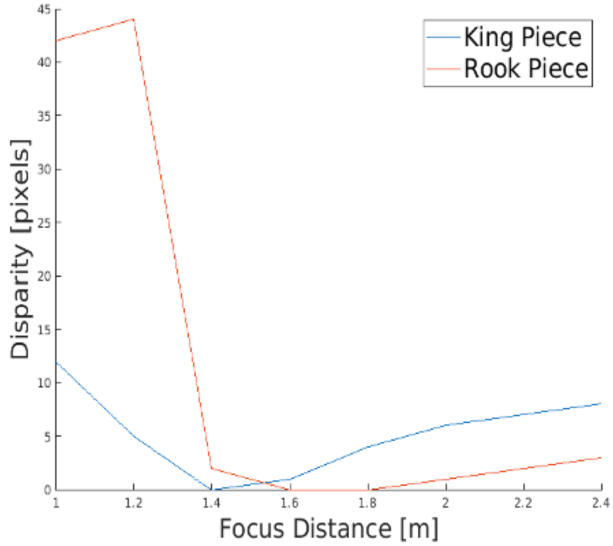

Figure 15: Disparity estimation as a function of focus distance for the Rook and King pieces. As can be seen the algorithm has succeeded to identify the correct object focal plane for the King piece, but found the same minimum disparity value for two adjacent focal distances (1.6 and 1.8 meters) for the Rook piece.

Figure 16 depict the mean magnitude gradient edges (sharpness estimation) as a function of focus distance for both the Rook and King pieces. Here we show the results for both pieces (King and Rook), but focuses on the result obtained for the Rook piece for the 1.6 and 1.8 meters focus distances. As can be seen the sharpness estimation for the 1.8 meter is greater than the 1.6 meter, which mean the algorithm chose it as the best focus for the Rook piece. This choice corresponds to the correct focus location (sharper image) of Rook piece. One should notice that the correct focus location could have been determined straight from the sharpness results. Due to lack of time, we did not explore this more into depth and it will have to be done as a part of the future work.

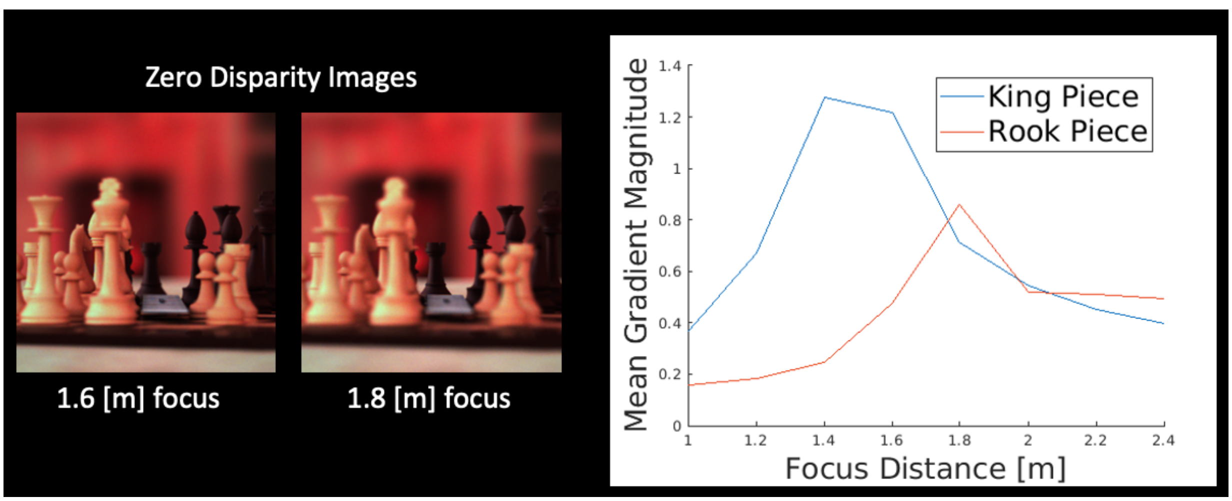

Figure 16: Mean magnitude gradient edges (sharpness estimation) as a function of focus distance for the two pieces (King & Rook). As can be seen, the sharpness estimation for the Rook piece for 1.8 meters is greater than 1.6 meters focus distances, which correspond to the correct focus distance (sharper image) of Rook piece

Conclusions:

In this work we demonstrated the dual pixel auto-focusing algorithm on two simulated 3D scenes (Slanted Bar and Chess Set) in ISET3D. Our main contribution is a simple graphical user interface (GUI) script that will be added to the ISET3D repository, and will enable future users to learn about the dual pixel auto-focusing algorithm. The dual pixel auto-focusing algorithm is part of the Phase Detection family auto-focusing algorithms, where focusing different objects in the scene can be achieved by estimating the disparity between the objects in the two forming sub-view images. The small the disparity the more in focus (sharper) the object in the image his. As we saw in the result section, our algorithm succeeded in the task to put different objects in the scenes in focus.

In general, the algorithm performances can be affected by many factors. For example, our implementation used the Zero Mean Normalize Cross Correlation (ZNCC) as a disparity estimator. This algorithm can not achieved sub-pixel (up to a certain degree) disparity resolution, so one potential future direction could be testing different disparity estimators. This is important because in this project we choose our optics to result in small depth of filed. This minimize the ambiguity (same minimum disparity values for two or more adjacent object focal planes), and enables better demonstration of the dual pixel auto-focusing algorithm. In many cases, for example smart phone cameras optics, the depth of filed is much larger, and better disparity estimation is necessary in order to put different objects in the scene in the correct focus.

In addition, the 3D scenes were captured in "ideal conditions", with high SNR and high contrast (in most of the cases). As we discussed, the ZNCC can be very sensitive to low contrast and noise, so another future direction could be exploring the affects of these parameters on our algorithm. Furthermore, as we saw in the last part of the report, a simple sharpness analysis yield the correct focus locations (for Rook and King pieces), so another interesting direction could exploring the performances of purely image processing methods, and where do they break. We suspect that these type of methods will perform poorly in different real life scenarios, because if that would not have been the case, other methods (like phase detection) would not have been developed.

In addition, our algorithm relies on the user choice of a region of interest (ROI) to focus on. As we can assume, different users will choose ROIs of the same object differently, which might affect the algorithms performances. As a result, it will interesting to test the error in focus estimation as a function of different ROIs around different objects. Finally, our algorithm was tested on a single simulated 3D real scene (Chess Set), so one important step will be testing it on a larger dataset of simulated scenes.

References:

1 Israni D., Patel S., Shah A. (2016) Comparison of Different Techniques of Camera Autofocusing. In: Satapathy S., Das S. (eds) Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 1. Smart Innovation, Systems and Technologies, vol 50. Springer, Cham. https://doi.org/10.1007/978-3-319-30933-0_14

2 M. Hamada, "Imaging device including phase detection pixels arranged to perform capturing and to detect phase difference," US20130088621, 2013.

3 M. G. Gluskin, R. M. Velarde, and J. Lee, "Phase detection autofocus noise reduction," US9420164, 2016.

4 Alba A., Aguilar-Ponce R.M., Vigueras-Gómez J.F., Arce-Santana E. (2013) Phase Correlation Based Image Alignment with Subpixel Accuracy. In: Batyrshin I., González Mendoza M. (eds) Advances in Artificial Intelligence. MICAI 2012. Lecture Notes in Computer Science, vol 7629. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37807-2_15

5 R. Garg, N. Wadhwa, S. Ansari, and J. T. Barron, "Learning single camera depth estimation using dual-pixels," in International Conference on Computer Vision, 2019.

6 A. Punnappurath, A. Abuolaim, M. Afifi and M. S. Brown, "Modeling Defocus-Disparity in Dual-Pixel Sensors," *2020 IEEE International Conference on Computational Photography (ICCP)*, 2020, pp. 1-12, doi: 10.1109/ICCP48838.2020.9105278.

7 C. D. Kuglin, "The Phase Correlation Image Alignment Method," in Proc. Int. Conf. Cybernetics and Society, 1975, pp. 163–165.

8 Bouguet, Jean-Yves. "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm." *Intel corporation* 5.1-10 (2001): 4.

9 Liang, Zhengfa, et al. "Learning for disparity estimation through feature constancy." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

Acknowledgments:

I would like to thank the course instructors: Dr. Brian Wandell, Dr. Joyce Farrell, David Cardinal and TA Zheng Lyu for all the help throughout the quarter.

I especially want to thank Thomas Goossens for his wonderful mentorship and guidance, and for spending hours with me on this project during the quarter.

Appendix:

The script and code for this project can be found on: https://github.com/ISET/iset3d/tree/master/tutorials/camera