Simulated depth of field

Ryan Ramsdell, Devesh Sullivan, and Amanda Sarsha

The aim of our application is to simulate a shallow depth of field in post production. Camera systems embedded in many consumer electronics like cell phones are generally smaller than traditional options. This reduced size in optics and sensor size puts limitations of the effective depth of field achievable, in result many of these systems are designed with a wider depth of field. Depth of field is critical in the artistic value of an image as it can be used as a powerful compositional tool in highlighting the subject matter or conversely obscuring unwanted parts of the image scene.

In attacking this problem we will initially restrict ourselves to an image domain with human subjects. The rationale being that selfies and portraits make for a large share of images taken today. With an increasingly connected society the issue of having a strong and reputable social media presence has become even more common. Our intention is to address the matter of resource access inequality (i.e., having a quality camera or ability to hire professional photographers) by implementing a solution that requires minimal hardware or effort in the image capture phase but still produces quality results comparable to professional images. Additionally, in implementing a software solution to this problem we eliminate the need for costly hardware systems.

The methodology can be broken into two distinct steps: image depth map calculation and rendering the final image.

Several solutions have been proposed in generating an image depth map. Depth from focus[1] and depth from defocus[2] both rely on the ability to alter the focus of the system, either through manipulation of relative distance between the lens and the image sensor or by adjustments to the aperture. They are further limited by requiring more than one image in order to calculate depth. Achieving good depth-resolution needs even more images, although Nayar and Nakawaga[3] suggest the use of Gaussian interpolation algorithms are sufficient to artificially increase resolution. Another approach is calculating depth from a pair of stereo images by examining disparity in matched scene points.[4] Longuet-Higgins assumes a known bijective correspondence between scene points in his methodology. Luckily this has been addressed by a multitude of parties, Scharstein and Szeliski[5] explore several algorithms that produce *dense* disparity maps, in that each pixel has a corresponding disparity estimate. At the time of publishing, they suggested the simple-shiftable-window algorithm presented by Hirshchmüller[6] was the most efficient algorithm. They maintain a taxonomy

1. S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16, no. 8 (August 1994): 824–831, ISSN: 0162-8828, doi:10.1109/34.308479.

2. Paolo Favaro, Andrea Mennucci, and Stefano Soatto, "Observing Shape from Defocused Images," *International Journal of Computer Vision* 52, no. 1 (April 2003): 25–43, ISSN: 1573-1405, doi:10.1023/A:1022366408068, https://doi.org/10.1023/A:1022366408068.

3. Nayar and Nakagawa, "Shape from focus."

4. H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature* 293 (September 1981): 133 EP -, https://doi.org/10.1038/293133a0.

5. Daniel Scharstein and Richard Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision* 47, no. 1 (April 2002): 7–42, ISSN: 1573-1405, doi:10.1023/A:1014573219977, https://doi.org/10.1023/A:1014573219977.

6. Heiko Hirschmüller, Peter R. Innocent, and Jon Garibaldi, "Real-Time Correlation-Based Stereo Vision

of currently best ranked algorithms online. Among the top ranked is a continuous 3D label stereo matching method using local expansion moves proposed by Taniai, Matsushita, Sato, and Naeumura.[7] This proposed method relies on an iterative approach with graph cuts used to propagate small changes in a localized area (3x3 cell). It is very slick in its approach but relies on Markov random fields and is outside of our scope to be implemented in time. Other approaches include image segmentation and dual-pixel autofocus systems operating on single images, both explored by Wadhwa et al..[8] Image segmentation generally makes assumptions about the scene trying to be captured and can use neural networks or other methods to separate common subject matter like people. The dual-pixel autofocus relies on specific sensor hardware and is not adaptable to most situations. Another segmentation technique proposed by Battiato et al.[9] performs color based image segmentation, then labels areas based on predetermined classifications gathered by examining common indoor and outdoor scenes and finally combines this with geometric properties to estimate the depth map.

Blurring and image composition present additional design decisions. As an aside, Wadhwa et al[10] give a naïve but accurate approach where each scene point projects a translucent disk (a scatter convolution of sorts) in the image with depths stacked as appropriate. This method is eschewed by the authors because it is computationally expensive on mobile phones and other low powered devices. Instead they offer a gather convolution coupled with a faster scatter convolution over a circle rather than a disk (filled circle) and a few other optimizations to make it run as quickly as possibly while still producing acceptable results. Other research has been done to create a circular separable convolution such using Gaussians like those produced by Garcia[11] or Niemitalo.[12] These approximate a perfectly circular blur disk produced by a perfectly circular aperture in a real lens and are very fast to compute. The näive approach suggested by Wadhwa[13] is more extensible in that it is easier to produce non-circular blur disks such as the hexagonal ones produced by physical camera systems with six-blade apertures.

Wholesale solutions to the problem also exist and are actively being researched by major phone and camera software producers. In particular, the portrait mode implemented by Google in their camera app for Android phones utilizes artificial intelligence to generate

with Reduced Border Errors," *International Journal of Computer Vision* 47, no. 1 (April 2002): 229–246, ISSN: 1573-1405, doi:10.1023/A:1014554110407, https://doi.org/10.1023/A:1014554110407.

7. Tatsunori Taniai et al., "Continuous Stereo Matching using Local Expansion Moves," *CoRR* abs/1603.08328 (2016), arXiv: 1603.08328, http://arxiv.org/abs/1603.08328.

8. Neal Wadhwa et al., "Synthetic Depth-of-field with a Single-camera Mobile Phone," *ACM Trans. Graph.* (New York, NY, USA) 37, no. 4 (July 2018): 64:1–64:13, ISSN: 0730-0301, doi:10.1145/3197517.3201329, http://doi.acm.org/10.1145/3197517.3201329.

9. Sebastiano Battiato et al., "Depth map generation by image classification," *Proc.SPIE* 5302 (2004): 5302-5302-10, doi:10.1117/12.526634.

10. Wadhwa et al., "Synthetic Depth-of-field with a Single-camera Mobile Phone."

11. Kleber Garcia, "Circular Separable Convolution Depth of Field," in *ACM SIGGRAPH 2017 Talks*, SIGGRAPH '17 (Los Angeles, California: ACM, 2017), 16:1–16:2, ISBN: 978-1-4503-5008-2, doi:10.1145/3084363.3085022, http://doi.acm.org/10.1145/3084363.3085022.

12. Ollie Niemitalo, *Circularly symmetric convolution and lens blur*, 2010, htpp://yehar.com/blog/?p=1495.

13. Wadhwa et al., "Synthetic Depth-of-field with a Single-camera Mobile Phone."

depth maps and create false depth of field.[14] They use a Structure From Motion technique with Markov random field inference methods to calculate the depth map and rendering was done by using the thin lens approximation. Apple has a similar product.

Our approach will use the segmentation technique given by Wadhwa et al.[15] as it is most adaptable to any imaging system which aligns with our general goal of being a hardware independent system. We are researching the ability to use Haar feature search and either mean-shift or graph based image segmentation to separate a human subject from the background of an image. The Haar algorithm would find the human subject and we intend to adjust the segmentation parameters so that the face we've recognized stays in one segment. After segmentation we will rely on reasonable assumptions about image composition that should more in most cases. Namely that the human subject will be in the foreground and that all other parts will be behind the subject. In addition, the if we have segments that are at different vertical heights along the image, the ones closer to the top will be assumed to be further from the camera as is true with a distant horizon or sky. Segmentation is limited in how many objects it can discern so we will also use Gaussian interpolation as suggested by Nayar et al.[16] as a way of increasing the depth-resolution. If time permits we may explore additional depth-map calculation techniques like depth from focus/defocus with images gathered from short video segments with changing focus or implementing an algorithm to gather depth from stereo.

We also plan to implement the more photorealistic image rendering technique from Wadhwa et al.[17] over the faster Gaussian approximations suggested by Garcia[18] and Niemitalo.[19] The current algorithm masks the image for each depth value and blurs with a circle of confusion proportional to the distance from the focal plane. We have restricted it to an only 8-bit depth-map which is expected to be fine given the limited resolution in our depth-map calculations. This keeps our render time down significantly. Focus is chosen manually at the moment, although it should be easy to automatically focus on the assumed human subject when our depth map calculations have been fully implemented as it designed to specifically look for human faces. We have the additional hope of allowing for arbitrary aperture shape, which requires updating our blur kernel generation algorithm to produce other shapes. In this consideration we will need to decide between procedurally generated shapes or up/down scaling pre-made kernel assets. Preliminary results can be seen in figure 1.

Currently we are using computer generated images to run tests in a controlled setting. With this we have the advantage of producing images of scenes with perfectly accurate and precise depth map calculations for which we can compare our own depth map against, as well as fine tuning the image composition section using said map. A future framework we wish to test our depth map algorithm are the datasets available from the taxonomy maintained

14. Carlos Hernández, *Lens Blur in the New Google Camera App*, April 2014, `https://ai.googleblog.com/2014/04/lens-blur-in-new-google-camera-app.html`.

15. Wadhwa et al., "Synthetic Depth-of-field with a Single-camera Mobile Phone."

16. Nayar and Nakagawa, "Shape from focus."

17. Wadhwa et al., "Synthetic Depth-of-field with a Single-camera Mobile Phone."

18. Garcia, "Circular Separable Convolution Depth of Field."

19. Niemitalo, *Circularly symmetric convolution and lens blur.*

on the Middlebury College website.[20][21][22][23][24] These tests will provide us with quantatative data to evaluate the effectiveness of our images. Furthermore we will test the accuracy of our segmentation and rendering phases on real-life stills by taking multiple stills of a scene with varying depth of field on a digital single lens reflex camera and comparing the results of our solution applied to the wide depth of field shots with the shallower optical results.

Moving forward, we still have to implement the depth-map calculation. Ideally this will be done by mid-November so that we can thoroughly test our application and then begin working on supplementary parts and fine-tuning the overall product.

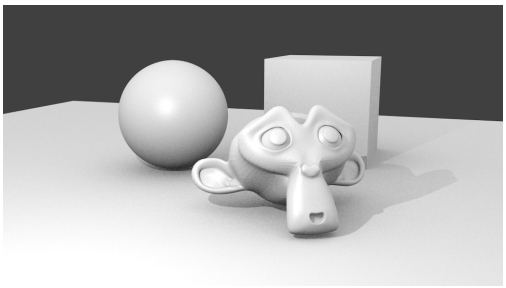
20. Scharstein and Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms."

21. D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.* Vol. 1 (June 2003), I–I, doi:10.1109/CVPR.2003.1211354.

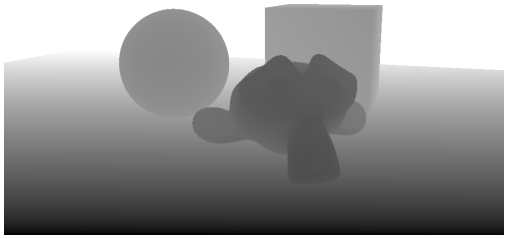
22. D. Scharstein and C. Pal, "Learning Conditional Random Fields for Stereo," in *2007 IEEE Conference on Computer Vision and Pattern Recognition* (June 2007), 1–8, doi:10.1109/CVPR.2007.383191.

23. Daniel Scharstein et al., "High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth," in *GCPR* (2014).

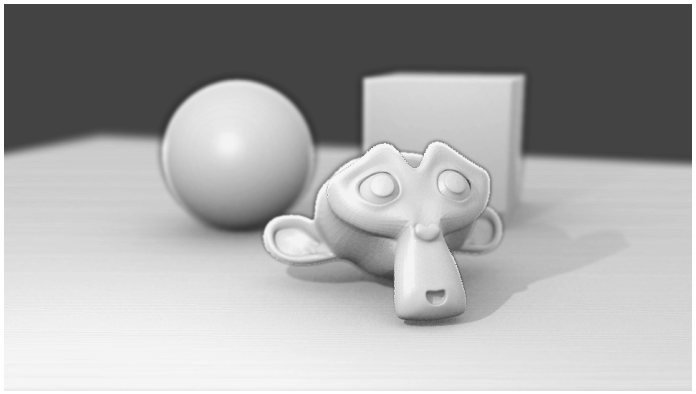
24. Heiko Hirschmüller and Daniel Scharstein, "Evaluation of Cost Functions for Stereo Matching," *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, 1–8.



(a) Original image.



(b) Image depth map.



(c) Final composited imaged.

(d) Original image.

(e) Image depth map.

(f) Final composited imaged.

Figure 1: Simulated depth of focus by blurring pixels based on depth.

References

Battiato, Sebastiano, Salvatore Curti, Marco La Cascia, Marcello Tortora, and Emiliano Scordato. "Depth map generation by image classification." *Proc.SPIE* 5302 (2004): 5302-5302-10. doi:10.1117/12.526634.

Favaro, Paolo, Andrea Mennucci, and Stefano Soatto. "Observing Shape from Defocused Images." *International Journal of Computer Vision* 52, no. 1 (April 2003): 25–43. ISSN: 1573-1405. doi:10.1023/A:1022366408068. https://doi.org/10.1023/A:1022366408068.

Garcia, Kleber. "Circular Separable Convolution Depth of Field." In *ACM SIGGRAPH 2017 Talks*, 16:1–16:2. SIGGRAPH '17. Los Angeles, California: ACM, 2017. ISBN: 978-1-4503-5008-2. doi:10.1145/3084363.3085022. http://doi.acm.org/10.1145/3084363.3085022.

Hernández, Carlos. *Lens Blur in the New Google Camera App*, April 2014. https://ai.googleblog.com/2014/04/lens-blur-in-new-google-camera-app.html.

Hirschmüller, Heiko, Peter R. Innocent, and Jon Garibaldi. "Real-Time Correlation-Based Stereo Vision with Reduced Border Errors." *International Journal of Computer Vision* 47, no. 1 (April 2002): 229–246. ISSN: 1573-1405. doi:10.1023/A:1014554110407. https://doi.org/10.1023/A:1014554110407.

Hirschmüller, Heiko, and Daniel Scharstein. "Evaluation of Cost Functions for Stereo Matching." *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, 1–8.

Longuet-Higgins, H. C. "A computer algorithm for reconstructing a scene from two projections." *Nature* 293 (September 1981): 133 EP -. https://doi.org/10.1038/293133a0.

Nayar, S. K., and Y. Nakagawa. "Shape from focus." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16, no. 8 (August 1994): 824–831. ISSN: 0162-8828. doi:10.1109/34.308479.

Niemitalo, Ollie. *Circularly symmetric convolution and lens blur*, 2010. htpp://yehar.com/blog/?p=1495.

Scharstein, D., and C. Pal. "Learning Conditional Random Fields for Stereo." In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. June 2007. doi:10.1109/CVPR.2007.383191.

Scharstein, D., and R. Szeliski. "High-accuracy stereo depth maps using structured light." In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.* 1:I–I. June 2003. doi:10.1109/CVPR.2003.1211354.

Scharstein, Daniel, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nesic, Xi Wang, and Porter Westling. "High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth." In *GCPR*. 2014.

Scharstein, Daniel, and Richard Szeliski. "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms." *International Journal of Computer Vision* 47, no. 1 (April 2002): 7–42. ISSN: 1573-1405. doi:10.1023/A:1014573219977. https://doi.org/10.1023/A:1014573219977.

Taniai, Tatsunori, Yasuyuki Matsushita, Yoichi Sato, and Takeshi Naemura. "Continuous Stereo Matching using Local Expansion Moves." *CoRR* abs/1603.08328 (2016). arXiv: 1603.08328. http://arxiv.org/abs/1603.08328.

Wadhwa, Neal, Rahul Garg, David E. Jacobs, Bryan E. Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan T. Barron, Yael Pritch, and Marc Levoy. "Synthetic Depth-of-field with a Single-camera Mobile Phone." *ACM Trans. Graph.* (New York, NY, USA) 37, no. 4 (July 2018): 64:1–64:13. ISSN: 0730-0301. doi:10.1145/3197517.3201329. http://doi.acm.org/10.1145/3197517.3201329.