Fast Bilateral-Space Stereo for Synthetic Defocus

Jonathan T. Barron[1], Andrew Adams[1], YiChang Shih[2], Carlos Hernández[1]
[1]Google [2]MIT

Given a stereo pair it is possible to recover a depth map and use that depth to render a synthetically defocused image. Though stereo algorithms are well-studied, rarely are those algorithms considered solely in the context of producing these defocused renderings. In this paper we present a technique for efficiently producing disparity maps using a novel optimization framework in which inference is performed in "bilateral-space". Our approach produces higher-quality "defocus" results than other stereo algorithms while also being $10-100\times$ faster than comparable techniques.

Most stereo algorithms work by assigning a disparity label to each pixel in an image. The core idea of our technique is that we avoid per-pixel inference by leveraging techniques for fast bilateral filtering [1, 2] to "resample" a dense stereo problem from pixel-space into a much smaller "bilateral-space". Bilateral-space is a resampling of pixel-space (or any vector space in which bilateral affinities are computed) such that small, simple blurs between adjacent vertices in bilateral-space are equivalent to large, edge-aware blurs in pixel-space. Instead of performing inference with respect to all pixels in an image, we instead solve a per-pixel optimization problem in this reduced bilateral-space. Our bilateral-space is defined such that it is cheap to compute, such that general optimization problems can be efficiently embedded into that space, and such that per-pixel depth-labelings we produce after inference have the edge-aware properties that make them useful for our defocus task. Because inference is done in this compact "bilateral-space" instead of pixel-space, our approach is fast and scalable despite solving a global optimization problem with non-local smoothness priors.

[1] Andrew Adams, Jongmin Baek, and Myers Abraham Davis. Fast high-dimensional filtering using the permutohedral lattice. *Eurographics*, 2010.

[2] Jiawen Chen, Sylvain Paris, and Frédo Durand. Real-time edge-aware image processing with the bilateral grid. *SIGGRAPH*, 2007.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

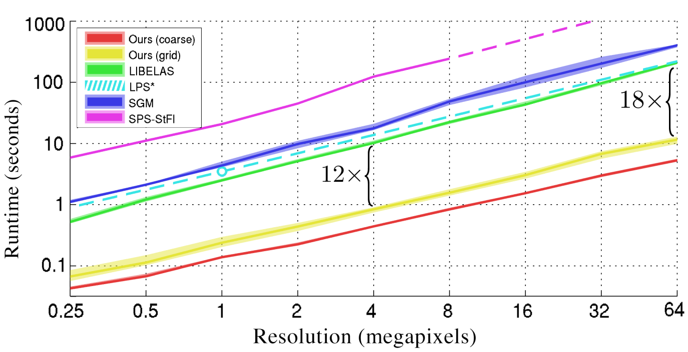
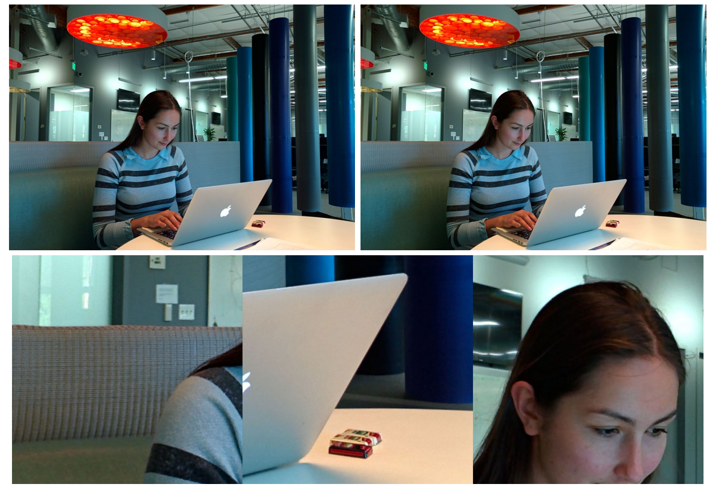


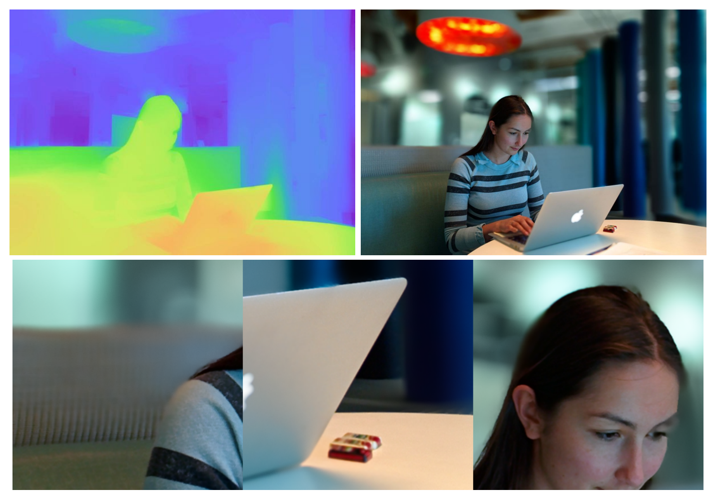
Figure 1: The runtimes for our algorithm and baselines, shown as a function of image resolution. Our algorithm is faster than all baselines by more than an order of magnitude.

Method	User 1	2	3	4	5	6	mean ± std.
Our Algorithm	175	185	162	134	165	145	161.0 ± 18.8
LIBELAS	9	5	10	8	16	8	9.3 ± 3.7
SGM	19	9	18	27	15	21	18.2 ± 6.0
CostFilter	15	12	13	24	23	11	16.3 ± 5.7
SPS–StFl	22	31	29	35	19	54	31.7 ± 12.4
SAD–lattice	24	22	32	36	26	25	27.5 ± 5.4

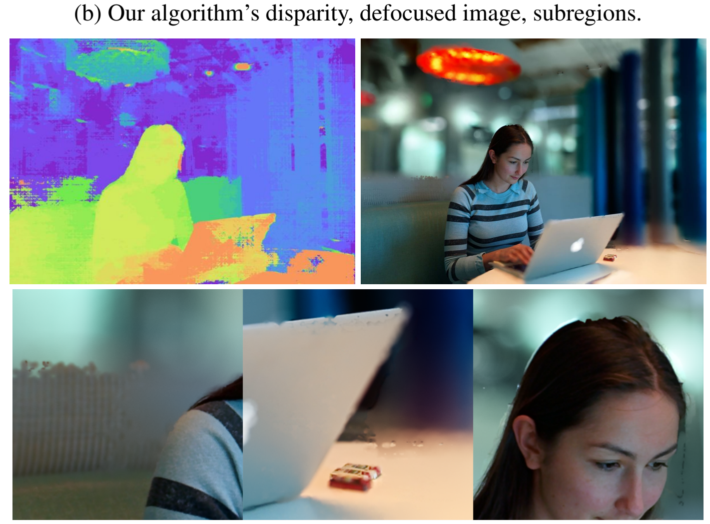
Table 1: The results of a user study, indicating user preference for the defocused renderings produced using different stereo algorithms. Images produced with our technique are preferred by a large margin.



(a) Input stereo pair with some cropped subregions of the right image.



(b) Our algorithm's disparity, defocused image, subregions.



(c) Semi-Global Matching's disparity, defocused image, subregions

Figure 2: Given a 4-megapixel stereo pair (2a), we produce a disparity map and then render a synthetically defocused image (2b). Images produced using other stereo algorithms (2c) tend to have unattractive artifacts.