Depth from Focus with Your Mobile Phone

Supasorn Suwajanakorn[1], Carlos Hernandez[2], Steven M. Seitz[1,2]
[1]University of Washington [2]Google Inc.

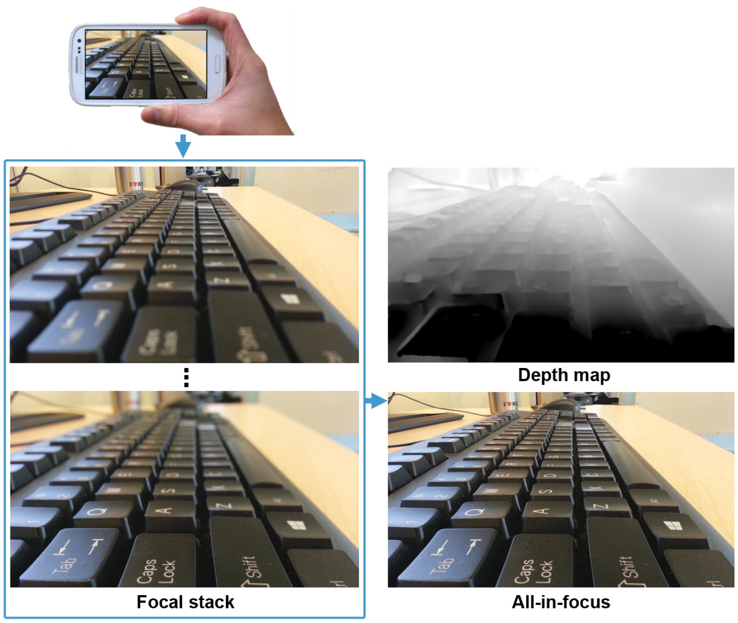

Figure 1: We compute depth and all-in-focus images from the focal stack that mobile phones capture each time you take a photo.

While prior depth from focus and defocus techniques operated on laboratory scenes, we introduce the first depth from focus (DfF) method capable of handling images from mobile phones and other hand-held cameras. Achieving this goal requires solving a novel uncalibrated DfF problem and aligning the frames to account for scene parallax. Our approach is demonstrated on a range of challenging cases and produces high quality results.

Every time you take a photo with your mobile phone, your camera rapidly sweeps the focal plane through the scene to find the best auto-focus setting. The resulting set of images, called a *focal stack*, could in principle be used to compute scene depth, yielding **a depth map for every photo you take**. While depth-from-focus (DfF) techniques have been studied for a couple decades, they have been relegated to laboratory scenes; no one has ever demonstrated an approach that works on standard mobile phones, or other hand-held consumer cameras. This paper presents the first successful demonstration of this capability, which we call *hand-held DfF*.

Hand-held DfF is challenging for two reasons. First, while almost all DfF methods require calibrated capture, supporting commodity mobile phones requires working in an uncalibrated setting. We must solve for the focal settings as part of the estimation process. Second, capturing a focal sweep with a hand-held camera inevitably produces motion parallax. The parallax is significant, as the camera motion is typically on the order of the aperture size (which is very small on a cell phone). I.e., the parallax is often larger than the defocus effect (bokeh radius).

One way to solve the uncalibrated DfF problem is to jointly solve for all unknowns, i.e., all camera intrinsics, scene depth and radiance, and the camera motion. The resulting minimization turns out to be intractable and one would need a good initialization near the convex basin of the global minimum for such non-linear optimization. In our case, the availability of the entire focal stack, as opposed to two frames usually assumed in depth-from-defocus problem, enables a relatively simple estimation scheme for the scene radiance. Thus, we propose a technique that first aligns every frame to a single reference and produces an all-in-focus photo as an approximation to the scene radiance. With the scene radiance fixed and represented in a single view, the remaining camera parameters and scene depth can then be solved in a joint optimization that best reproduces the focal stack

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

Figure 2: Datasets captured with a hand-held Samsung Galaxy phone. From left to right (number of frames in parenthesis): plants(23), bottles(31), metals(33). Top row shows the all-in-focus stitch. Bottom row shows the reconstructed depth maps.

In addition, we propose a refinement scheme to improve depth map accuracy by incorporating spatial smoothness and an approach to correct the bleeding problem for saturated, highlight pixels, known as *bokeh*, in the estimation of an all-in-focus image.

We present depth map results and all-in-focus images in figure 2 and a refocusing application of our method in figure 3.

In conclusion, we introduced the first depth from focus (DfF) method capable of handling images from mobile phones and other hand-held cameras. We formulated a novel uncalibrated DfD problem and proposed a new focal stack aligning algorithm to account for scene parallax.

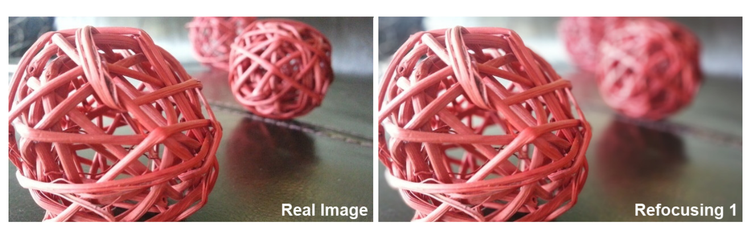

Figure 3: Left is a real image from a focal stack sequence. Right shows a synthetic rendering with increased aperture to amplify the depth-of-field effect and emphasize the foreground object.