Light Field Photography for CGI/Live Action Video Integration

Matt Lathrop, Keenan Molner

Department of Computer Science, Department of Electrical Engineering, Stanford University

Motivation

In the film industry today, many shots are composed of both real and virtual objects and composited together to form a single frame of video. One cheap and easy way of achieving compositing is using a green screen, but it only allows for a live action actor to be placed into a virtual world. The alternative, placing a virtual object into a live action scene can be much more complicated due to the lack of depth information in the live action scene.

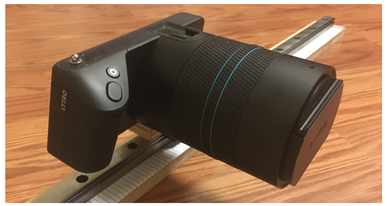
We propose the use of light field photography to eliminate this time consuming and expensive manual process and replace it with a computational process. If we can calculate the depth of each object in a live action shot, then it is trivial to figure out which pixels in the live action shot are occluding the virtual object, as well as which pixels lie behind the virtual object.



Our Techniques

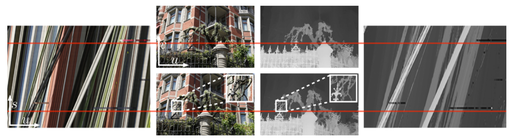
Image Capture

We constructed linear translation stage to mount a camera upon. This stage, 1m in length, allowed us to capture many images along the same axis, spaced apart by width of the camera sensor. These images are the individual frames of the live action film, for which we calculated a depth map at each frame.



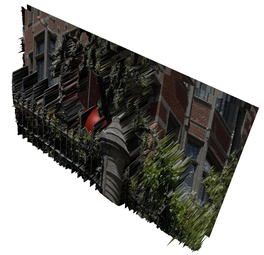
Depth Calculation

We implemented the Kim, et al. depth calculation algorithm in MATLAB. The algorithm searches for edges across high resolution cross-sectional slices of the image stack, calculates the slope of these edges, and calculates a depth proportional to the slope. This depth is then propagated across all frames where the edge is still detected before downsampling the image to a coarser resolution and repeating.



CGI Compositing

In a 3D compositing program, we precisely matched a virtual camera's movement to the real camera's movement. Then, we focused our virtual camera on a rectangle showing our live action video sequence, textured with the depth map of the corresponding video frame. Finally, we inserted a virtual object into the scene. Since the virtual and real cameras follow the same motion path, the virtual object seamlessly blends with the live action.



Related Work

The Fraunhofer Institute constructed a large camera array to capture lightfields. These images were then used to construct a depth map estimate. While their results were promising, their rig was large and their depth estimation based on a limited number a viewpoints within the array.

The Lytro cameras capture lightfields and run proprietary depth calculation algorithms that allows for depth map creation and image refocusing. A camera like this, but with a higher pixel density and faster readout times could be used to film a scene and add in composited effects, in post production. Unfortunately, in practice, we've found the calculated depth maps from the Lytro camera to be underwhelming and often make gross oversimplifications of occlusions and regions with high noise content.

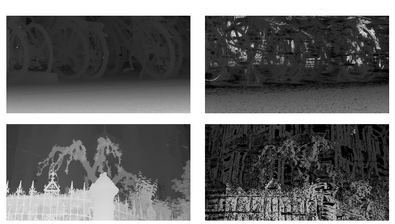
Disney Research (Kim, et al.) has published results on calculating high resolution depth maps with a single camera across a linear track. Their method looks at the edges of features of the scene extruded in cross-sectional slices of the lightfield image stack. The steeper the slope of an edge in the cross-sectional lightfield stack, the greater distance the feature is from objects focused at the center of the stack. Additionally, the Kim, et al. method starts by examining areas with highest spatial frequency and then explores larger areas in a fine-to-coarse manner, rather than many other image processing techniques that evaluate from coarse-to-fine, smoothing along the way and reducing the resolvability of fine details.

Results

Depth Calculations

Our algorithm successfully detected the fine details of the scenes and provided a nuanced depth map for high frequency components of the image stack. We struggled to fill in the areas of constant texture. We think this is likely caused by some simplifications, and assumptions made when we implemented the Kim, et al. paper.

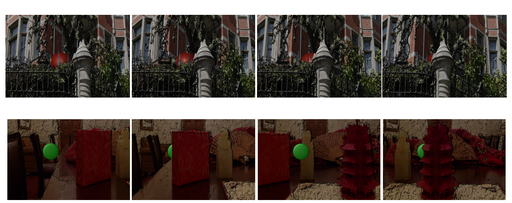
Kim, et al. Ours



CGI Compositing

We composited a 3D scene with the Kim et al. dataset and our own image capture with the Lytro camera.

The Kim et al. dataset allowed for more intricate compositing since depth information was available for smaller, more detailed regions.



Future Work

The Kim et al. depth map algorithm provides highly detailed depth maps, but we were unable to fully replicate their results and their speed claims. It would take upward of 30min to calculate the depths for 100 frames. Efforts to correct and parallelize the algorithm would allow for longer live action scenes.

Currently, 3D compositing programs require the depth map to be computed as a polygon texture, frame by frame, before overlaying the image to the shape. An improved pipeline for applying depth maps to image sequences in 3D compositing program would simplify the rendering pipeline.

References

Ziegler, Matthias, et al. "Multi-camera system for depth based visual effects and compositing." Proceedings of the 12th European Conference on Visual Media Production. ACM, 2015.

Zilly, Frederik, et al. "Light-field acquisition system allowing camera viewpoint and depth of field compositing in post-production." (2013): 5-2.

Kim, Changil, et al. "Scene reconstruction from high spatio-angular resolution light fields." ACM Trans. Graph. 32.4 (2013): 73-1.

N.d. Metro. Web. 8 Mar. 2016. <http://metro.co.uk/2014/12/18/from-hero-to-zero-famous-action-films-with-special-effects-trickery-removed-4991825/>.