EE367 Final Project proposal

Matt Lathrop and Keenan Molner

Abstract:

We propose a project to improve the simplicity and speed for filmmakers to incorporate captured video with effects and graphics composited on the computer. Currently, when an artist looks to insert composited graphics into a video captured with digital camera technology, they must match the camera motion of the computer graphic with the camera used to capture the live action, rotoscope and condition the live action scene to prepare it for a 3D insertion, which can take great effort, and finally composite the two video files together, often with lots of tweaking and correction.

We propose to use a light field camera to capture the live action and to implement a modern depth calculation algorithm to extract a depth map for each frame of the video file. With an accurate depth map of the live action scene imported to the 3D compositing program, we can now match our virtual camera's motion to match the live action camera's motion and place our 3D composited content directly into the live action at the correct depth. No longer will tedious masking and scaling be required to properly insert virtual content into live scenes, complete with occlusions and changing focal planes.

Motivation:

In the today's film industry many shots are composed of both real and virtual objects. The process of combining live action footage with computer generated images is called compositing. One way of achieving compositing is using a green screen. In this method the live action footage has a constant color screen in the background of the image that can be "keyed" out. This is an inexpensive and easy way of compositing 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. In modern films this is achieved by a manual process called rotoscoping. Essentially and artist goes through the live action footage and traces out certain objects that need to act as occluders to the virtual object. Our proposal is 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 image need to act as a background to your inserted virtual object and which ones need to act as an occluder.

Related Work:

Previous attempts at the goal of computationally finding depth maps for use in compositing have proved unsuccessful for one of two reasons: physically large camera rigs, and poor depth map reconstruction. Fraunhofer, attempted to solve this problem recently in their paper "Multi-camera system for depth based visual effects and compositing". They used an array of cameras along with a "hero" camera that used a beam splitter to allow both of the cameras to see the scene from the same

angle. While their results were promising, their rig was incredibly large and their depth estimation based on a limited number a viewpoints.

The Lytro camera group constructed a lightfield camera (essentially a multi-camera array, but with lower resolution due to multiplexing across the same sensor) and runs a proprietary depth calculation algorithm 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. 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. While the hardware is compact and predisposed for lightfield imaging, the software package and algorithm for depth calculation is too loose than what's required for film compositing.

A group from Disney Research has also published a paper on calculating high resolution depth maps with either one lightfield camera or a camera array. Their method is based on looking at the edges of features of the scene extruded by cross sectional slices of the light field image stack and calculating the slope of the the features as they change in perspective. The more change a pattern demonstrates across the (u,v) lightfield stack, the greater distance the feature is from objects focused at the center of the stack. Both an object that lies very close and very far from the center show great change in distance across the lightfield stack, but with opposite directions. By calculating the slope of these lines, the relative depth of one object from another can be extrapolated. Additionally, Disney's 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.

Project Overview:

In short, our project aims to demonstrate the ability of capturing a live action scene with a lightfield camera and to calculate a high resolution depth map for each frame of video. With this depth map, we aim to to properly composite a virtual object and a live action scene with no manual tracing.

First, we will capture a short video sequence with the Lytro Illum camera and a single dimensional time lapse motion controller, like the Radian from Alpine Labs. Upon image capture, we intend to run these images through a matlab implementation of the (u,v) slope to depth algorithm, presented in the Disney paper. While the algorithm will look at the entire stack of lightfield images, we will disregard color information and ability to refocus the image for the sake of calculating a high resolution, smooth, and accurate depth map for each frame of the captured video.

Once we have a good depth map for our image, we can now apply our depth map to a surface in Maya and then texture map our image onto this surface. In essence this should create a plane for each depth in the depth map containing the pixels at that depth in the live action image. This will allow us to view the image just as it was captured, but with the added ability to insert virtual objects between the various depth planes in the image.

If we are able to implement an efficient process for importing the depth map into Maya, then we intend to not just apply this algorithm to still images from the Lytro Illum but to make a short time-lapse video using the camera. In this case we would compute a per-frame depth map, and then update the surface in Maya for each frame. Using a time lapse camera movement rig, we can show how objects placed in front of our object will change whether they occlude the object as the camera changes its viewpoint.

Milestones:

- Talk to Lytro about software: 2/8
- Review Depth papers: 2/8 - 2/15
- Implement depth algorithm: 2/12 - 2/22
- Capture Light field data 2/22-2/29
- Apply algorithm and export for maya 2/22 - 2/29
- Put into Maya 2/29 - 3/4
- Make interesting demo 2/29 - 3/4
- Writeup 3/11

References:

- Disney depth paper: https://s3-us-west-1.amazonaws.com/disneyresearch/wp-content/uploads/20140725221413/Kim13.pdf
- Rotoscoping with depth maps
    - Matthias Ziegler, Andreas Engelhardt, Stefan Müller, Joachim Keinert, Frederik Zilly, Siegfried Foessel, and Katja Schmid. 2015. Multi-camera system for depth based visual effects and compositing. In *Proceedings of the 12th European Conference on Visual Media Production* (CVMP '15). ACM, New York, NY, USA, , Article 3 , 10 pages. DOI=http://dx.doi.org/10.1145/2824840.2824845
- Other depth detection paper - http://hci.iwr.uni-heidelberg.de/HCI/Research/LightField/papers/WG12_cvpr.pdf
- General light field video stuff