Real Time 3D Scene Reconstruction from Multiple Cameras

Alex Watras

February 13, 2016

1 Introduction

The goal of this project will be to implement algorithms to generate a 3d model of a scene from multiple video cameras in real time. The goal is to use a 3D depth map to generate a video feed by projecting the image to a new perspective. This will allow us to generate stereoscopic views of a scene, as well as allowing us to move through a 3d scene more naturally than with other methods. We hope that this method will allow us to have a 3D view of a remote location in near real time. Possible uses for this type of technology are first person navigation of remote vehicles, more immersive remote communication software, and a more natural viewing environment that surgeons could use during minimally invasive laparascopic surgeries.

2 Related Work

There are many existing methods for recreating a 3D scene from 2D images. Structured light methods such as the methods proposed by [2],[1], and the method used by the Microsoft kinect use a light source to project a pattern onto the scene. If we know the location of the light source relative to the camera, we can use basic triangulation techniques as outlined in [2] to determine a depth map of the scene.

Structure from motion algorithms such as are used in [3] do not need any external light source. Instead, they attempt to match points between multiple images and use those to estimate the relative position of the two cameras. Once those positions have been calculated, the same triangulation methods can be used to calculate the depth mapping of the scene.

On this project, we will attempt a method more similar to [4] where we can cut down on computation times by using a camera array with known relative pose to simplify the calculations necessary for 3D scene reconstruction.

3 Technical Description

Our goal is to allow user's to experience a remote location with the same visual sensory input as if they were there. In order to do that, we need to be able to understand how light is projected onto each of the cameras. In the case of our camera array the relative locations should always be the same. This means that once we have calibrated our cameras, we will be able to skip the pose estimation step and simplify calculations down to just the triangulation step. However, since we are not using structured light, we will need some other method of identifying whether two points in the images of separate cameras came from the same location. Once we can pair the pixels together, we match up the pixel rays and look for the intersection of corresponding pixel rays in 3D space. This will allow us to generate a depth map of the scene.

Once we have a depth map for the scene, We want to use that depth map to recreate a new view of the scene. By projecting the scene into two new viewpoints spaced appropriately, we should be able to generate a stereoscopic view of the scene.

4 Milestones

1. Given a camera array with known relative poses, calculate the corresponding camera fundamental matrices using a calibration algorithm.
2. Identify corresponding pixel pairs in images.
3. First we must transform the pixels of each camera into a vector showing the possible locations in 3D space from which the pixel could have come.
4. By intersecting the vectors of corresponding pixels in each camera's image, we can generate a depth map of the scene.
5. Given the camera matrices calculate a depth map of the scene.
6. From the depth map, generate a stereoscopic view from a desired viewpoint by projecting the scene points back through a new virtual camera.
7. Simplify computation so that on-line video transmission from the desired viewpoint is possible.
8. (Optional) Integrate with a display platform such as google cardboard to allow near real time exploration of a remote location via a virtual reality environment.

5 References

1. Gupta, M., & Nayar, S. K. (2012). Micro Phase Shifting Conventional Phase Shifting Projected images Micro Phase Shifting. IEEE Conference on Computer Vision and Pattern Recognition, 813820.
2. Lanman, D., & Taubin, G. (2009). Build Your Own 3D Scanner : 3D Photography for Beginners. ACM SIGGRAPH 2009 Courses, 94. http://doi.org/10.1145/1665817.1665819
3. Snavely, N., Seitz, S. M., & Szeliski, R. (2006). Photo tourism. ACM Transactions on Graphics, 25(3), 835. http://doi.org/10.1145/1141911.1141964
4. Yang, R., Welch, G., & Bishop, G. (2002). Real-time consensus-based scene reconstruction using commodity graphics hardware. Proceedings - Pacific Conference on Computer Graphics and Applications, 2002-Janua, 225234. http://doi.org/10.1109/PCCGA.2002.1167864