Improved Stereoscopic 3D Reconstruction Using Keypoint Detection

EE 368 Project Proposal

Harvey Han (hanhs@stanford.edu)
Ifueko Igbinedion (ifueko@stanford.edu)

Introduction

This project aims to improve on existing methods for 360° stereoscopic image reconstruction by utilizing multiple perspectives of panoramic images stitched from fisheye lenses. Based on translation input from a user wearing a head mounted display, we can use a combination of depth estimation, image warping and keypoint detection to reconstruct 3D views without excessive use of computationally expensive inpainting algorithms. By estimating which camera capture location maps to a user's translation, and matching image features close to image areas that need inpainting, we hope to achieve fast, if not real-time 3D stereoscopic reconstruction of real world scenes.

Figure 1: After detecting foreground and background components using a depth map, we encounter areas that were hidden in the original perspective. Because inpainting is costly and inaccurate, incorporating data from different perspectives can help fill in these holes much more quickly[1].

Camera Capture System and 3D Reconstruction

We aim to experiment with different multiples and configurations of 360° fish eye lens cameras. Code exists to convert two spherical viewpoints into one stereoscopic view, but not for a wide range of perspectives. By using 3 to 4 staggered viewpoints, we can capture spherical image data for a single scene at multiple viewing angles. Knowing the configuration of our camera capture system, we can then create mappings from the cross product of these camera locations and use these mappings to decide which two 360° views to use for the initial RGB-D depth map estimation[2]. We can then warp the resulting view and inpaint unknown background areas by interpolating with reconstructed views adjacent to the current camera, detecting alignment with keypoint detection and filling in areas that seem to be new background information close to unknown areas. Keypoint detection in this case entails calculating feature descriptors for each image, noting which features are the most prominent in each image. Using Scale Invariant Feature Transform (SIFT) descriptors[3], we can then compare and match features from a shifted view in the same local neighborhood as unknown areas of the original warped view. After these keypoints are detected, inpainting should become computationally inexpensive.

Applications and Challenges

Using this 360° stereoscopic image reconstruction, we can potentially develop some meaningful virtual reality applications. For example, one possible application aside from the typical virtual reality experience would be to allow users to blur and focus different depth locations of the reconstructed environment. The most significant challenge we expect to face is developing a fast data processing scheme for real time 3D stereoscopic translation. The existing reconstruction algorithms are computationally expensive and prohibit real-time results, so one of our final goals is to achieve (or give the illusion of achieving) real-time results.

References

[1] Zheng, K.C.; Sing Bing Kang; Cohen, M.F.; Szeliski, R., "Layered Depth Panoramas," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on* , vol., no., pp.1-8, 17-22 June 2007

[2] L. Zhang and Wa James Tam. Stereoscopic image generation based on depth images for 3d tv. Broadcasting, IEEE Transactions on, 51(2):191–199, June 2005.

[3] Lowe, David G. "Distinctive image features from scale-invariant keypoints."*International journal of computer vision* 60.2 (2004): 91-110.