EE 368 Final Project Proposal
Tyler S. Jordan, Skanda Shridhar with Jayant Thatte

Depth-based stereoscopic image rendering and 3D reconstruction has been an important area of research in multimedia, broadcasting and in computer vision. The area has received a lot of attention from the broadcast research community for its applications in 3D Television (3DTV) and Free Viewpoint Television (FTV) [1], [2], [3], [4]. Similarly, depth-based 3D perception is an important prerequisite for computer vision, robotics and augmented reality applications [5], [6], [7]. With the recent advances in omnidirectional camera rigs, computational methods for postproduction, and head-mounted display (HMD) technologies, VR is on its way to become the next cinematic medium. At the core of all of these applications is the ability to produce precise and accurate depth maps for the scene under consideration. The ability to generate reliable depth maps helps in synthesizing novel views, which is a key step in supporting these varied applications.

Humans have an innate ability to perceive depth from stereo imagery; however, conventional stereo correspondence algorithms are generally incapable of producing reliable dense disparity maps. In this project, we wish to investigate the use of machine learning techniques to tackle this problem. For our work, we plan to use the stereo dataset from The KITTI Vision Benchmark Suite [8], [9]. The dataset consists of roughly 200 training sequences and roughly 200 test sequences. Morphological features (edges, shapes, key-points etc.) will be extracted from the training images using image processing techniques. These will be the inputs to our machine learning algorithm which will detect the disparity at each point between the images and in turn estimate the depth at the corresponding pixels. We will explore convolutional neural networks as a candidate approach and support vector machines as an alternative. The produced disparity maps will then be passed on to a post-processing stage. The raw disparity will be denoised, smoothed and filtered to output the final estimated disparity. The output will be evaluated based on the fraction of pixels for which the estimated disparity differs from its ground-truth value by more than a predecided threshold.

Note: This work will be jointly submitted as a project for EE368 by "Skandavimal Shridhar" and "Tyler S. Jordan" and for CS229 with “Jayant Thatte”. Care will be taken to ensure that the image processing part of the project is clearly distinguishable from the machine learning part.

References:
[1] L. Zhang and W. J. Tam, “Stereoscopic image generation based on depth images for 3d tv,” Broadcasting, IEEE Transactions on, vol. 51, no. 2, pp. 191–199, June 2005.
[2] G. Thomas and O. Grau, “3d image sequence acquisition for tv & film production,” in 1st InternationalSymposium on 3D Data Processing Visualization and Transmission, 2002, pp. 320–326.
[3] J. Flack, P. Harman, and S. Fox, “Low bandwidth stereoscopic image encoding and transmission,” in Proceedings of SPIE-IS&T Electronic Imaging, 2003, pp. 206–214.
[4] C. Fehn, “Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv,” pp. 93–104, 2004.
[5] H. Kim and A. Hilton, “3d scene reconstruction from multiple spherical stereo pairs,” International Journal of Computer Vision, vol. 104, no. 1, pp. 94–116, 2013.
[6] H. Kim, M. Sarim, T. Takai, J.-Y. Guillemaut, and A. Hilton, “Dynamic 3d scene reconstruction in outdoor environments,” in 3DPVT, 2010.
[7] M. Schonbein and A. Geiger, “Omnidirectional 3d reconstruction in augmented manhattan worlds,” in Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on, Sept 2014, pp. 716–723.

[8] A.Geiger, P. Lenz and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite", CVPR, 2012.
[9] M. Menze and A. Geiger, "Object Scene Flow for Autonomous Vehicles", CVPR, 2015.