Comparing the robustness of different depth map algorithms using light fields

Linda Banh (lbanh@stanford.edu), Fang-Yu Lin (fangyuln@stanford.edu)

Goals

To compare different depth map algorithms using light fields and to see how it performs when reconstructing 3D scenes. Reconstruction quality (mean-squared error) and efficiency (computation time) will be taken into account during the comparison.

Methods

We plan to implement three different kinds of algorithms proposed in past research and compare them with a ground truth depth map (which will be taken from 4D Light Field Benchmark[1], a public light field and depth map dataset). To reconstruct the 3D scene, we will display our results in point clouds (scatter3) using MATLAB. In addition, if we have time, we will transfer the depth maps into Meshlab, an open source system to visualize 3D meshes, and render them to see the quality of the reconstruction compared to the ground truth and the 3D reconstruction results using the different algorithms.

The first algorithm we will implement is creating different focused images using a focal stack [2]. The image focused at specific depths will be constructed by using a shift and add algorithm on the focal stack [3]. We will then find the gradient of each image focused at each depth and find maximal gradient among the stack to build our depth map. This will be used as our baseline to compare with the two more advanced algorithms below.

There are two different advanced methods to be implemented. In the first algorithm, we will explore a method proposed by Disney Research where they use epipolar planes on densely sampled light fields [4] to calculate the gradient of the light field. Next, the gradients will be used to segment the foreground and background [5] to create their depth maps. In the second algorithm, we will explore a method proposed by Anisimov and Stricker [6], where they create depth maps based an existing idea of line fitting in a 4-dimensional light field space. Line fitting typically uses a linear regression solution to match pixels of an image with a correlated depth. However, this is computationally expensive. Therefore, Anisimov and Stricker propose using a semi-global matching algorithm for line fitting, a fast approximation of the gradient at each depth, rather than doing pixel by pixel like line fitting.

For all three algorithms, we will use the dataset from 4D Light Field Benchmark[1] since they provide a ground truth depth map for their rendered scene and this could be beneficial for us to compare our depth map extracted from all the algorithms mentioned above. To compare the maps quantitatively, we will use mean-squared error between each algorithm and the ground truth depth map and compare their computation time.

References

[1] Honauer, Katrin, et al. "A dataset and evaluation methodology for depth estimation on 4d light fields." Asian Conference on Computer Vision. Springer, Cham, 2016.

[2] “Focus Stacking.” *Wikipedia*, Wikimedia Foundation, 18 Nov. 2018, en.wikipedia.org/wiki/Focus_stacking.

[3] “Shift-and-Add.” *Wikipedia*, Wikimedia Foundation, 28 Aug. 2017, en.wikipedia.org/wiki/Shift-and-add.

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

[5] Yücer, Kaan, et al. "Efficient 3D object segmentation from densely sampled light fields with applications to 3D reconstruction." *ACM Transactions on Graphics (TOG)* 35.3 (2016): 22.

[6] Anisimov, Yuriy, and Didier Stricker. "Fast and efficient depth map estimation from light fields." *2017 International Conference on 3D Vision (3DV)*. IEEE, 2017.

Indication whether you will use an Android device

No