Efficient and Scalable Depthmap Fusion

Enliang Zheng
http://www.cs.unc.edu/~ezheng
Enrique Dunn
http://www.cs.unc.edu/~dunn
Rahul Raguram
http://www.cs.unc.edu/~rraguram
Jan-Michael Frahm
http://www.cs.unc.edu/~jmf

Department of Computer Science
University of North Carolina
Chapel Hill, NC USA

The estimation of a complete 3D model from a set of depthmaps is a data intensive task aimed at mitigating measurement noise in the input data by leveraging the inherent redundancy in overlapping multi-view observations. In this paper we propose an efficient depthmap fusion approach that reduces the memory complexity associated with volumetric scene representations. By virtue of reducing the memory footprint we are able to process an increased reconstruction volume with greater spatial resolution. Our approach also improves upon state of the art fusion techniques by approaching the problem in an incremental online setting instead of batch mode processing. In this way, are able to handle an arbitrary number of input images at high pixel resolution and facilitate a streaming 3D processing pipeline.

Our proposal builds upon recently proposed heightmap representations [2, 3] and introduces a wavelet based compression mechanism for every column in the heightmap in order to attain a reduced parametric representation of each column. The heightmap fusion method takes a set of depthmaps as input, along with external and internal camera parameters. From the external camera parameters, we can determine a ground plane for the scene [4], which serves as the lower $x-y$ boundary of the fusion volume. Following this, the $x-y$-plane (ground plane) is partitioned into cells, where the x, y size of the cell matches the desired resolution of the 3D-model. For each cell we can define a column representing the volume above the cell. One of the computational advantages of the original heightmap fusion algorithm, in contrast to earlier volumetric fusion methods, is that the fusion is solved independently within each column, allowing for easy parallelization of the fusion process. There are two main observations that we leverage: (a) in the basic heightmap formulation, columns of the 3D volume are processed independently of each other and (b) when considering a single column , it is typically the case that the occupancy function function values change smoothly within a segment, with sharp transitions between adjacent segments. These sharp transitions correspond to the change from occupied to empty space in the physical world. Accordingly, wavelet based compression techniques are very well suited for modeling such volumetric data, as the transitional intervals of the input data signal can be accurately represented using a small number of wavelet coefficients [5].

The goal of our multi-layer heightmap estimation module is to transform the input occupancy function values for all the voxels belonging to a given column into an output binary segmented set of occupancy layers. We model the associated segmentation problem as a *directed acyclic graph* traversal problem and efficiently find the optimal solution through dynamic programming (DP). Our multi-layer approach differs from [3] in the following:

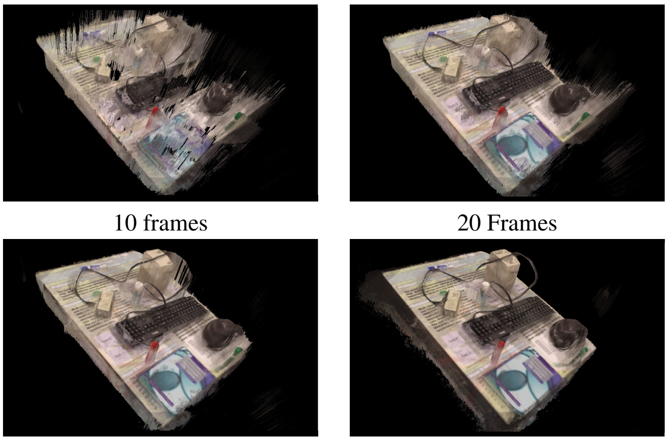
1. We deploy a general graph structure that is not restricted in regard to the parity of the number of layers nor in the topology of the assigned layers. In [3] strong assumptions on the observed scene needed to be made to define the graph structure.

2. We perform the dynamic programming in a reduced search space. We identify and quantify segments of homogeneous occupancy classification and use them as atomic elements to be labeled by the DP procedure. The importance of this pre-processing is that it enables more efficient processing at finer voxel resolutions while mitigating the memory requirements of DP search.

In our experiment, camera poses and depthmaps are computed using the pipeline of [1]. We compared our depthmap fusion approach against the one proposed in [3]. Fig. 1(a) (b) illustrate an improved robustness to noise provided by our method. We additionally tested our fusion method with the output of a SfM pipeline for close range scene reconstruction. We utilized a 16MP DSLR camera to obtain a total of 100 images (we name this data set randObject). Figure 2 depicts the sequential process of incrementally building the 3D model through online depthmap fusion. Our depthmap fusion runs at 33 Hz on one NVIDIA Tesla C2050/C2070 graphics card. Computing a 9-layer heightmap on a volume of 100x100x100 given one image takes 24.9 ms, within which the procedure of compression and decompression only takes 1.6 ms.

(a) Our Results (b) Results from [3]

Figure 1: (a) and (b) are comparative results for crowd sourced data. Both images show 3D models of the Brandenburg Gate.



10 frames 20 Frames

40 Frames 100 Frames

Figure 2: Incremental Online Depthmap Fusion. The 3D model is improved as more images are added.

[1] Jan-Michael Frahm, Pierre Fite-Georgel, David Gallup, Tim Johnson, Rahul Raguram, Changchang Wu, Yi-Hung Jen, Enrique Dunn, Brian Clipp, Svetlana Lazebnik, and Marc Pollefeys. Building rome on a cloudless day. In *ECCV*, 2010.

[2] David Gallup, Jan-Michael Frahm, and Marc Pollefeys. A heightmap model for efficient 3d reconstruction from street-level video. In *3DPVT*, 2010.

[3] David Gallup, Marc Pollefeys, and Jan-Michael Frahm. 3d reconstruction using an n-layer heightmap. In *DAGM*, 2010.

[4] Richard Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Computer Vision*, 2: 1–104, 2006.

[5] Wikipedia. Wavelet transform, 2011. URL "http://en.wikipedia.org//Wavelet_transform".