Module 4 - Week 4 - Optional task 3

Hicham El Muhandiz[1][0000−1111−2222−3333], Julia Ariadna Blanco Arnaus[1][0000−1111−2222−3333], Marcos Muñoz González[1][0000−1111−2222−3333], Rosana Valero Martínez[1][0000−1111−2222−3333]

Universitat Autònoma de Barcelona, Campus UAB, Edifici O, 08193 Cerdanyola del Vallès, Barcelona

1 Depth map fusion review

In this section we do a comparison between the two methods used in the papers *A Volumetric Method for Building Complex Models from Range Images* [1] and *RoutedFusion: Learning Real-time Depth Map Fusion* [2] to bridge the concepts from novel and traditional applications of fusion of depth maps.

In the first paper [1], Brian Curless and Marc Levoy present the classical approach to fusing noise depth maps, which is to average truncated signed distance functions (TSDF). In particular, they present a volumetric method for integrating range images with features like uncertainty representation, utilization of all data, incremental updating, efficiency, robustness against outliers, no topological restrictions, and ability to fill holes. The objective of this method is to combine multiple range images to produce smooth, highly detailed models consisting of up to 2.6 million triangles.

The algorithm employs a continuous implicit function, D(x), which is the weighted signed distance of each point x to the nearest range surface along the line of sight to the sensor.

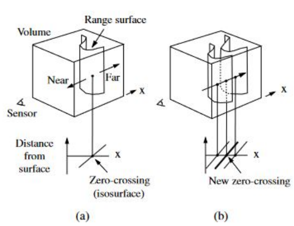

Fig. 1: Unweighted signed distance functions in 3D.

For each voxel, a cumulative signed distance function, D(x), and a cumulative weight W(x) are obtained by applying the combination rules that the authors introduced in this paper. Expressed as an incremental calculation,

$$D_{i+1}(x) = \frac{W_i(x)D_i(x) + \omega_{i+1}(x)d_{i+1}(x)}{W_{i+1}(x) + \omega_{i+1}(x)} \tag{1}$$

$$W_{i+1}(x) = W_i(x) + \omega_{i+1}(x) \tag{2}$$

where, d_i (x) and w_i (x) are the signed distance and weight functions from the ith range image.

The algorithm starts by resetting all voxel weights to zero, allowing new data to overwrite the initial grid values. The range images are divided into triangles and transformed into a mesh with weights at each vertex. The voxel grid is then updated by calculating the signed distance, by casting a ray from the sensor through each voxel and intersecting it with the triangle mesh, and the weight, by linearly interpolating vertex weights. After that, equations 1 and 2 can be applied. The zero-crossing surface can be extracted from the volumetric grid at any point during range image merging.

This method has advantages like local (truncated) updates, efficient fusion of noisy measurements, and low computational cost. However, it also has limitations like linear updates requiring minimum thickness assumption, inability to handle gross outliers, and need for pre-filtering and tuning of parameters for specific scenes and sensors. Thickening artifacts and vanishing surfaces can occur along edges and for thin objects.

In paper [2], the aim is to overcome the limitations of the previous study. To do so, instead of a simple linear fusion of depth information, they propose a neural network that predicts non-linear updates to better account for typical fusion errors. Their network consists of a 2D depth routing network and a 3D depth fusion network, which effectively deals with sensor-specific noise and outliers. This is particularly useful for surface edges and thin objects, where the original method tends to result in thickening artifacts.

Their pipeline consists in: First, a Depth Routing Network pre-processes depth maps to denoise and correct outliers before passing them to the Depth Fusion Network. After that, instead of processing each ray independently as in standard TSDF fusion, The Depth Fusion Network extracts a local camera-aligned voxel grid with TSDF data, weighted via trilinear interpolation from the corresponding global voxel grids, and then computes the local TSDF update v_t*. This predicted update v_t* is transferred back into the global coordinate frame to get v_t, and then, integrated into the global TSDF volumes, V_t, W_t, using the equations 3 and 4, which are the update equations introduced by Curless and Levoy:

$$V_t(x) = \frac{W_{t-1}(x)V_{t-1}(x) + \omega_t(x)v_t(x)}{W_{t-1}(x) + \omega_t(x)} \tag{3}$$

$$W_t(x) = W_{t-1}(x) + \omega_t(x) \tag{4}$$

where v_t and w_t are the signed distance update and its corresponding weight at time step t.

The results of their method demonstrate smoother flat surfaces in comparison to standard TSDF fusion. Additionally, the method provides improved reconstruction of thin structures with less surface thickening artifacts. In Figure 2, they provide evidence that their method significantly reduces noise and eliminates thickening effects compared to traditional TSDF fusion.

In summary, both papers aim for 3D reconstruction of a scene from 2D data using a volumetric representation as a voxel grid and depth sensor data. They both integrate multiple views to obtain a more complete and accurate 3D representation. However, they differ in their methods, input data, output and processing speed:

Fig. 2: Qualitative results on ShapeNet test data between TSDF (right) and Routed Fusion (left) methods

- Approach: [1] uses a volumetric representation to build complex models from range images while [2] uses a machine learning approach to real-time depth map fusion.
- Input data: [1] takes in range images while [2] takes in multiple depth maps as input.
- Output: [1] outputs a complex 3D model of the scene, while [2] outputs a single fused depth map.
- Processing: [1] processes range images offline, while [2] processes depth maps in real-time.
- Performance: [1] focuses on the accuracy and detail of the resulting 3D models, while [2] focuses on real-time performance.

Overall, we conclude that the use of NN-based function approximators allows for real-time processing and can improve the accuracy and efficiency of the fusion process compared to traditional methods.

References

1. B. Curless and M. Levoy. A Volumetric Method for Building Complex Models from Range Images. In Proc. SIGGRAPH, 1996.
2. RoutedFusion: Learning Real-time Depth Map Fusion Silvan Weder, Johannes L. Schönberger, Marc Pollefeys, Martin R. Oswald May 2020.