Deep 3D Reconstruction

Daniel G. Aliaga

February 2025

Well-used (non-DL) Packages

- COLMAP: https://colmap.github.io/
  - SFM and MVS pipeline (from UNC)

- Bundler, PMVS/PMVS2, CMVS, VSFM: https://www.di.ens.fr/cmvs/
  - SFM and MVS pipeline (from UW originally)

Some datasets

- SunCG:
  - https://sscnet.cs.princeton.edu/
- Pano3D:
  - https://vcl3d.github.io/Pano3D/

Image-based 3D Object Reconstruction: State-of-the-Art and Trends in the Deep Learning Era

- https://arxiv.org/pdf/1906.06543.pdf

Plan3D: Viewpoint and Trajectory Optimization for Aerial Multi-View Stereo Reconstruction

- https://ait.ethz.ch/plan3d
- https://files.ait.ethz.ch/projects/plan3d/VideoPlan3d.mp4

- Method: perform a first capture and reconstruction (using COLMAP), analyze and determine best path for a new capture that explores least observed areas, then re-capture and reconstruct

Multi-Robot Collaborative Dense Scene Reconstruction

- https://kevinkaixu.net/projects/multirobot.html

  (see video on webpage)

- Method: coordinate several robots using a 360-degree scanner

ScanComplete: Large-Scale Scene Completion and Semantic Segmentation for 3D Scans

- https://arxiv.org/pdf/1712.10215.pdf
- https://www.youtube.com/watch?v=5s5s8iH0NF8
- Method:
  - for a given partial input scan, they infer missing geometry and predict semantic labels on a per-voxel basis (e.g., 1480×1230×64 voxels (≈ 70×60×3m))
  - Train network to take in a partial truncated distance field (TDF) and output complete and labelled; uses 5400 scenes for training and 150 for testing from SUNCG
  - CNN based model

TSDF (Truncated Signed Distance Function)

- Popular in some scanning works

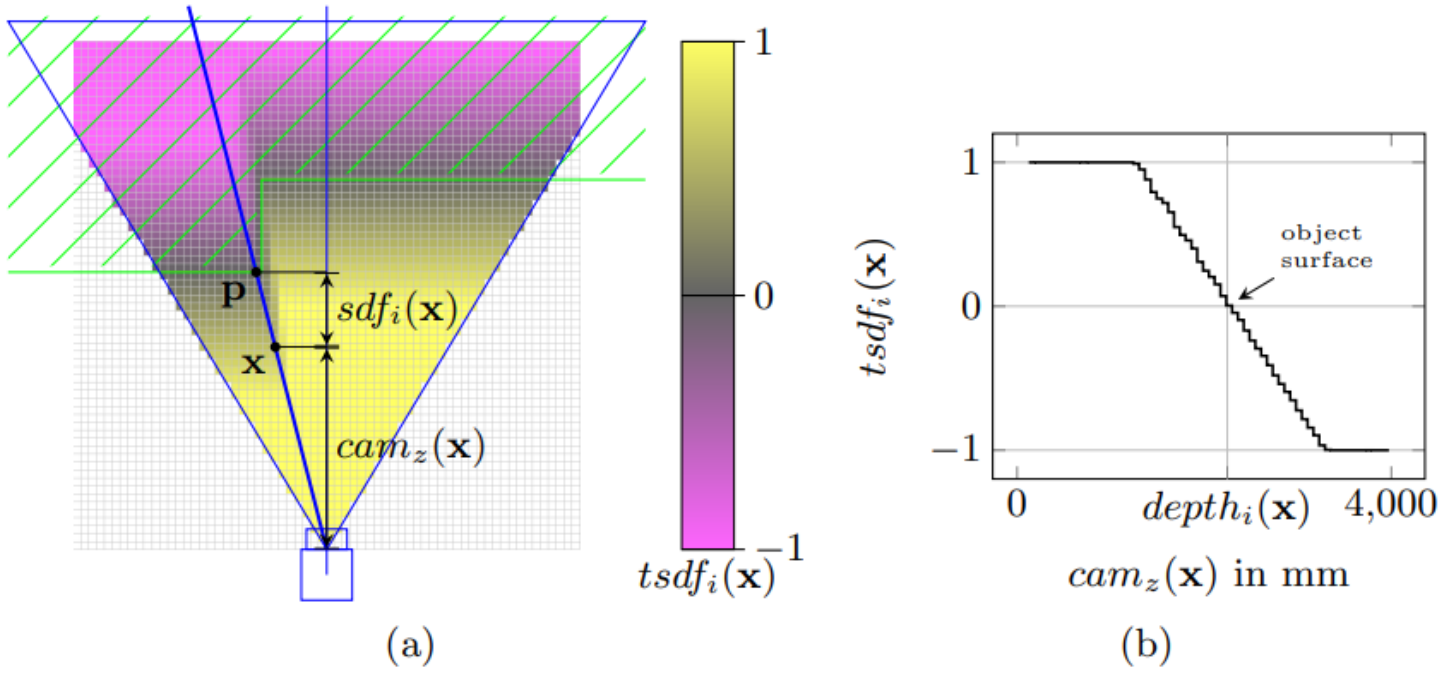


Fig. 1. 2D TSDF example. (a) Solid object (green), camera with field of view, optical axis and ray (blue), and TSDF grid (unseen voxels are white, for others see color bar). The signed distance value of voxel $\mathbf{x}$ is determined by the depth of the corresponding surface point $\mathbf{p}$ and the voxel's camera distance $cam_z(\mathbf{x})$. b) 1D TSDF sampled along the ray through $\mathbf{p}$ with $t = 1000$ mm. Object surface is at zero crossing.

ScanComplete: Large-Scale Scene Completion and Semantic Segmentation for 3D Scans

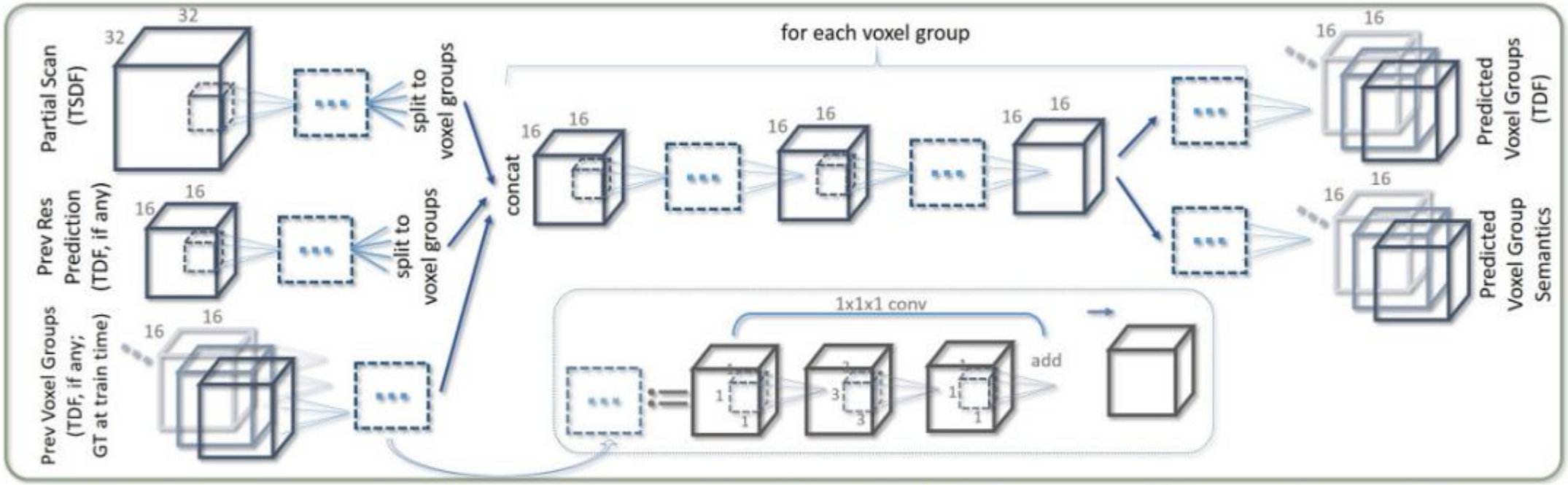


Figure 3. Our ScanComplete network architecture for a single hierarchy level. We take as input a TSDF partial scan, and autoregressively predict both the completed geometry and semantic segmentation. Our network trains for all eight voxel groups in parallel, as we use ground truth for previous voxel groups at train time. In addition to input from the current hierarchy level, the network takes the predictions (TDF and semantics) from the previous level (i.e., next coarser resolution as input), if available; cf. Fig. 1.

VolumeFusion: Deep Depth Fusion for 3D Scene Reconstruction

- https://openaccess.thecvf.com/content/ICCV2021/papers/Choe_VolumeFusion_Deep_Depth_Fusion_for_3D_Scene_Reconstruction_ICCV_2021_paper.pdf

Method	2D Depth Evaluation				3D Geometry Evaluation			
	AbsRel	AbsDiff	SqRel	RMSE	$\mathcal{L}_1$	Acc	Comp	F-score
COLMAP [37]	.137	.264	.138	.502	.599	.069	.135	**.558**
MVDepthNet [39]	.098	.191	.061	.293	.518	.040	.240	.329
GPMVS [20]	.130	.239	.339	.472	.475	**.031**	.879	.304
DPSNet [22]	.087	.158	.035	.232	.421	.045	.284	.344
Murez *et al.* [32]	.061	.120	.042	.248	.162	.065	.130	.499
VolumeFusion (ours)	**.049**	**.084**	**.021**	**.164**	**.141**	.038	**.125**	.508

Table 1. **Quantitative results on ScanNet dataset [9].** We provide two metrics: depth evaluation and 3D geometry evaluation.

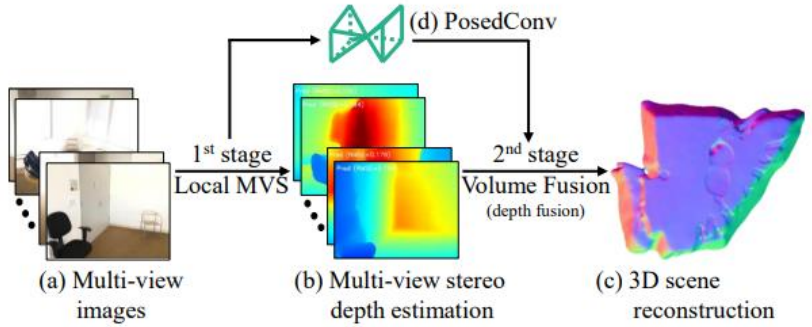


Figure 1. **Volume fusion.** Given (a) multi-view images and their camera parameters, our network aims at 3D scene reconstruction. (b) First, we estimate local multi-view depth maps. (b) Second, we introduce differentiable depth fusion with the guidance of pose-invariant features from (d) our PosedConv.

- Method: first SFM/MVS, then 3D CNN called PosedConv
- https://www.youtube.com/watch?v=NMB9_dch4BI

PlanIT: Planning and Instantiating Indoor Scenes with Relation Graph and Spatial Prior Networks

- https://www.youtube.com/watch?v=qgYY-oCPV3I&feature=youtu.be
- https://drive.google.com/file/d/1CJCM6EQyeUWwxdk6tl8cVxEIhV7s3DoA/view
- Method:
  - This is the opposite goal --- generate indoor scenes
  - Given a room, generate plausible content based on a partial graph
  - Train a to produce a graph (uses CNN/FCN), uses part of SUNCG (15,000 of 40,000 scenes)
  - Newer version of: “Fast and Flexible Indoor Scene Synthesis via Deep Convolutional Generative Models”, https://arxiv.org/pdf/1811.12463.pdf

Deep Reconstruction: RGBD

Deep Depth Completion of a Single RGB-D Image

- https://deepcompletion.cs.princeton.edu/

Learning View Priors for Single-view 3D Reconstruction

- http://hiroharu-kato.com/projects_en/view_prior_learning.html
- https://arxiv.org/abs/1811.10719

Transformable Bottleneck Networks

- https://kyleolsz.github.io/TB-Networks/

(includes video)

- Also:

https://www.youtube.com/watch?v=zem03fZWLrQ

Starting at 11:10…

Foray in Pixels and Single Images…

Fast Separation of Direct and Global Images Using High Frequency Illumination

- http://www1.cs.columbia.edu/CAVE/publications/pdfs/Krishnan_TOG06.pdf

HDR From Single Image

- http://www1.cs.columbia.edu/CAVE/publications/pdfs/Nayar_CVPR00_2.pdf

Learning Single Camera Depth Estimation using Dual-Pixels

- https://arxiv.org/abs/1904.05822

https://www.youtube.com/watch?v=zem03fZWLrQ

Starting at beginning…

Shapes and Points…

ShapeNet

- https://www.shapenet.org/

- *Also*
  - *A Point Set Generation Network for 3D Object Reconstruction from a Single Image,* CVPR 2017
    - https://arxiv.org/pdf/1612.00603.pdf

PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation

- https://www.youtube.com/watch?v=Cge-hot0Oc0
  - (start around middle)
- https://arxiv.org/pdf/1612.00593.pdf
- Also:
  - PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space
    - https://arxiv.org/pdf/1706.02413.pdf

ScanNet

- http://www.scan-net.org/

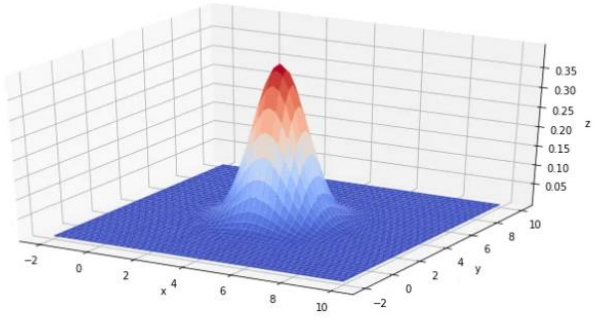
Neural Rendering…

Neural Point-Based Graphics

(before 3D Gaussians)

- https://dmitryulyanov.github.io/neural_point_based_graphics
- https://arxiv.org/pdf/1906.08240.pdf

Recent 3D Gaussian Splatting



- https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/
- https://3dgstutorial.github.io/3dv_part1.pdf

Image-guided Neural Object Rendering

- https://niessnerlab.org/projects/thies2020ignor.html

Special Cases…

Reconstructing Scenes with Mirror and Glass Surfaces

- https://www.gcc.tu-darmstadt.de/home/proj/mirror_reconstruction/mirror_reconstruction.en.jsp
- http://download.hrz.tu-darmstadt.de/media/FB20/GCC/project_files/mirror_reconstruction/MirrorReconstruction.mp4

InverseFaceNet: Deep Monocular Inverse Face Rendering

- https://niessnerlab.org/projects/kim2018inversefacenet.html