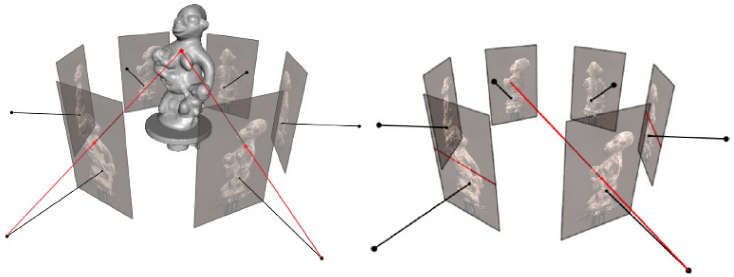
2021 ICCV OCTOBER 11-17 VIRTUAL

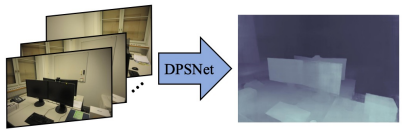
A Confidence-based Iterative Solver of Depths and Surface Normals for Deep Multi-view Stereo

Wang Zhao*, Shaohui Liu*, Yi Wei, Hengkai Guo, Yong-Jin Liu

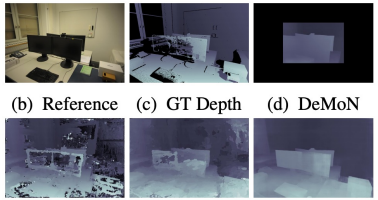
Multi-view Stereo



Images COLMAP

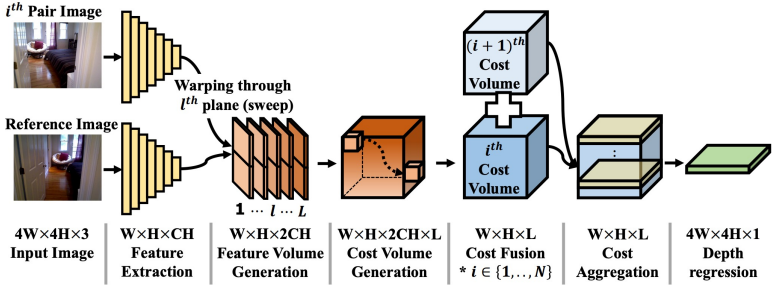


(a) Input unstructured image sets, output depth map.



(b) Reference (c) GT Depth (d) DeMoN

(e) COLMAP (f) DeepMVS (g) Ours



Advantage: complete

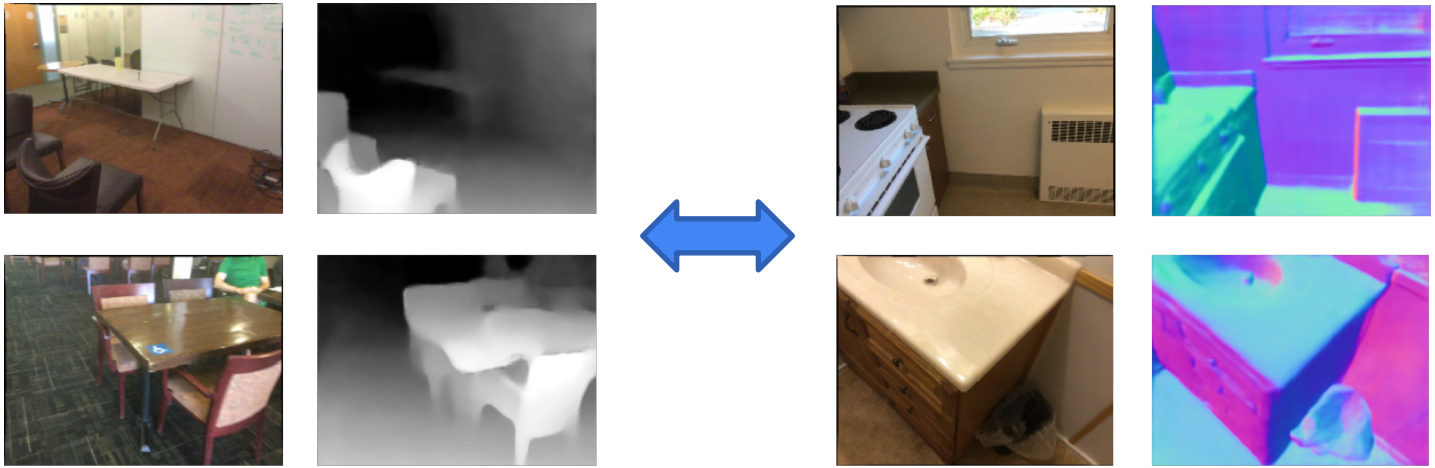
Disadvantage: higher depth error and lower quality 3D structure compared to conventional MVS !

Furukawa et al., Multi-View Stereo: A Tutorial
Schonberger et al., Pixelwise View Selection for Unstructured Multi-View Stereo, ECCV 2016

Im et al., DPSNet: End-to-end Deep Plane Sweep Stereo, ICLR 2019

Depth and Surface Normal

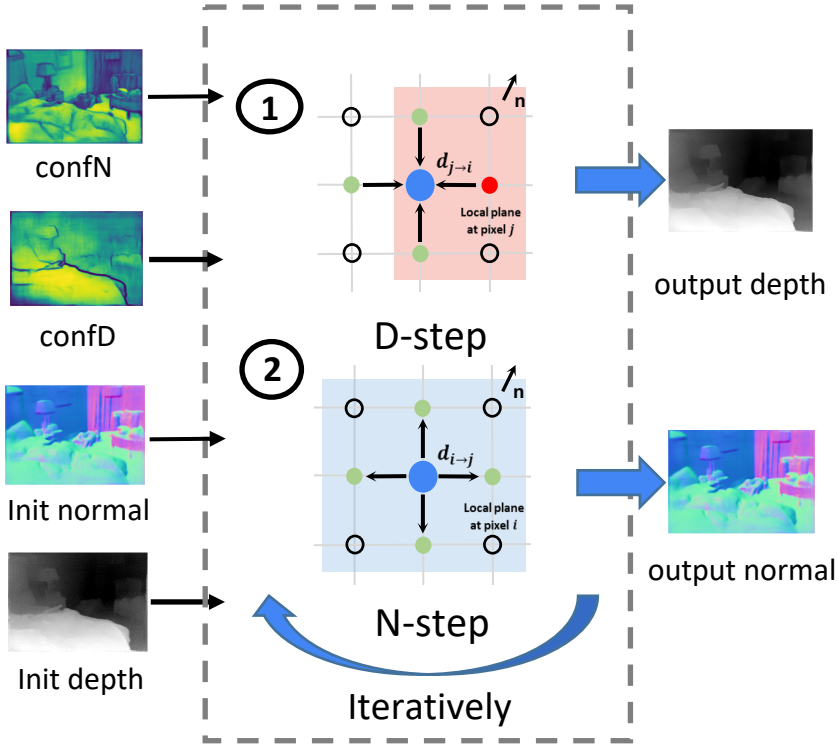
- Surface normal predictions are usually more robust on low-texture (often plane) regions.
- Depth and surface normal could benefit each other in joint training, as shown in GeoNet, NAS, CNM.



Our motivation: explicit modeling of planar structure within DeepMVS!

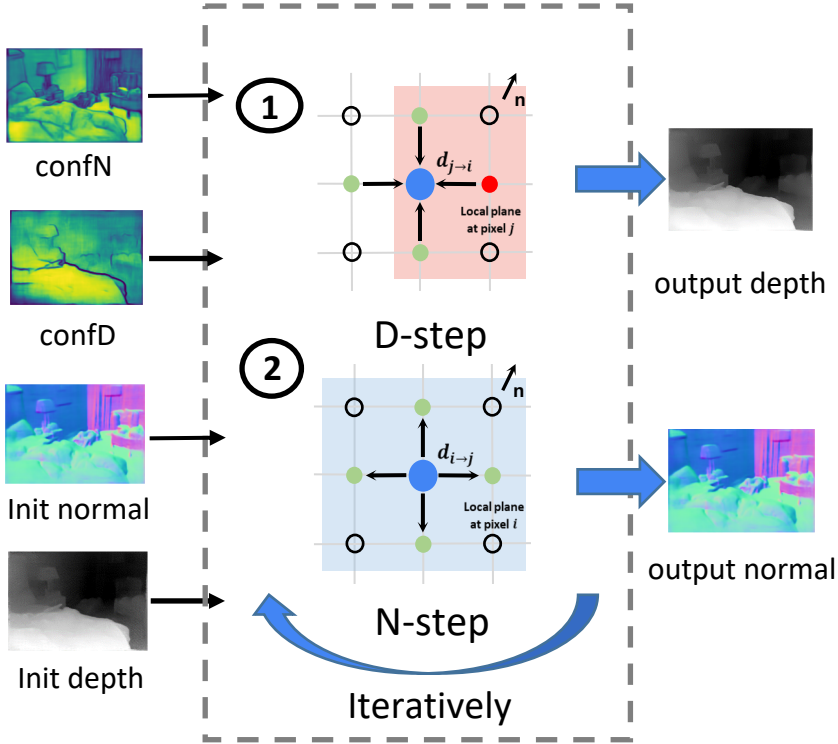
Qi et al., GeoNet: Geometric Neural Network for Joint Depth and Surface Normal Estimation, CVPR 2018
Kusupati et al., Normal Assisted Stereo Depth Estimation, CVPR 2020
Long et al., Occlusion-aware Depth Estimation with Adaptive Normal Constraints, ECCV 2020

Confidence-based Iterative Depth-Normal Solver

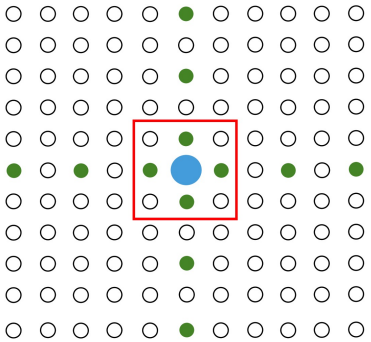


- Energy potential construction based on locally planar assumption
- Iterative optimization over depth and surface normal subproblems
- Closed-form solutions (thus differentiable) for both subproblems

Confidence-based Iterative Depth-Normal Solver

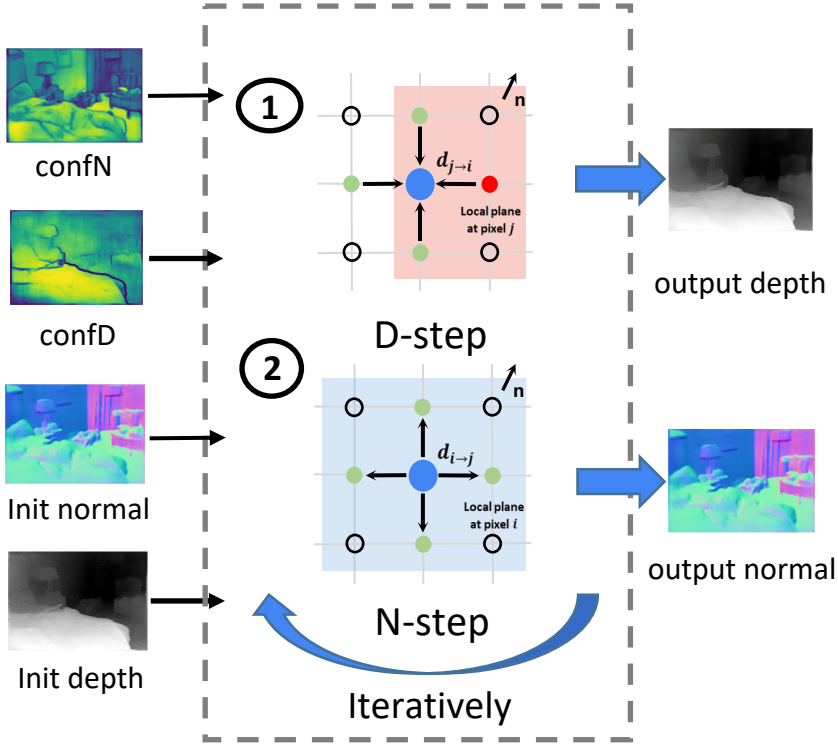


- Energy potential construction based on locally planar assumption
  - $E_{total} = \alpha E_{data} + E_{plane}$
  - Local neighbors for each pixel determined by predefined checkerboard



(a) Checkerboard

Confidence-based Iterative Depth-Normal Solver



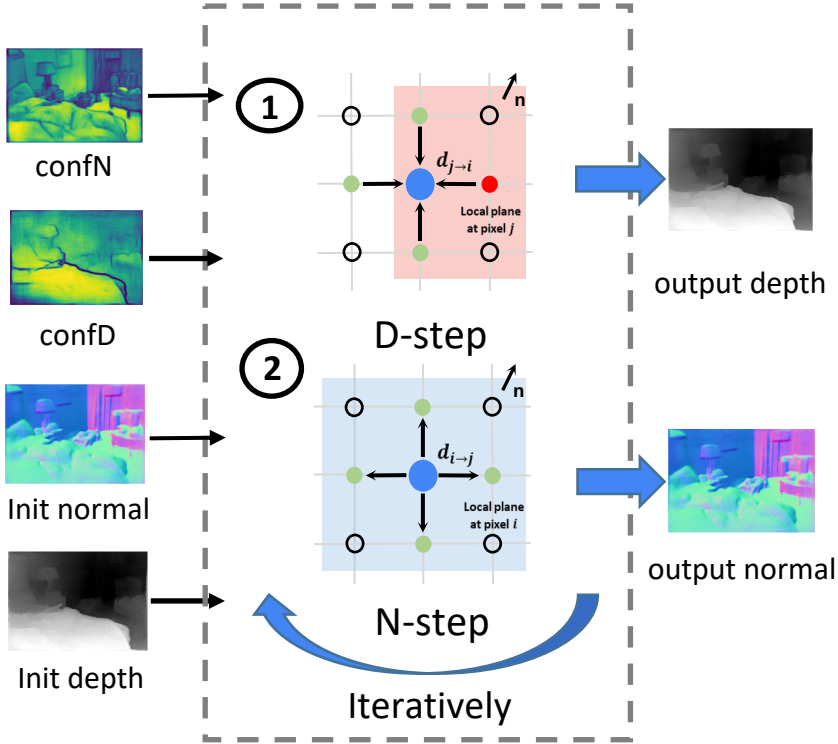
- D-Step: fix normal and update depth

$$\min_d E_{total} = \min_d E_d$$

$$E_d = \alpha \sum_i c_i (d_i - \hat{d}_i)^2 + \sum_i \sum_{j \in N(i)} c_j w_{ij} (d_i - d_{j \to i})^2$$

$$d_i^* = \frac{\alpha c_i \hat{d}_i + \sum_{j \in N(i)} c_i w_{ij} d_{j \to i}}{\alpha c_i + \sum_{j \in N(i)} c_i w_{ij}}$$

Confidence-based Iterative Depth-Normal Solver



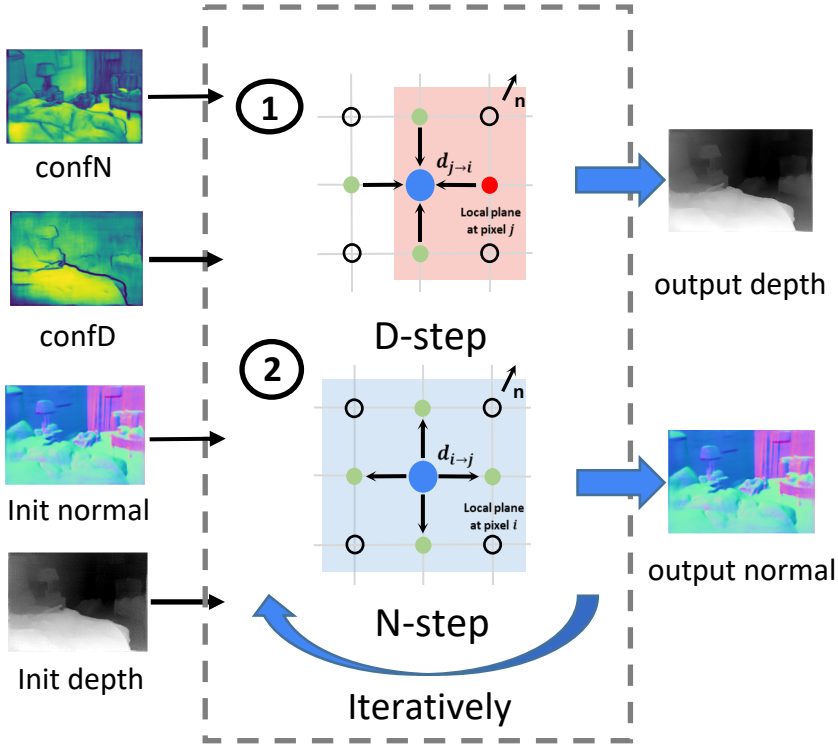
- N-step: fix depth and update normal

$$\min_n E_{total} = \min_n E_n$$

$$E_n = \alpha \sum_i c_i ||n_i - \hat{n}_i||^2 + \sum_i \sum_{j \in N(i)} c_j w_{ij} D_n(d_j, P(x_i, d_i, n_i))$$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} a_i^* \\ b_i^* \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

Confidence-based Iterative Depth-Normal Solver



- Energy potential construction based on locally planar assumption
- Iterative optimization of depth and surface normal subproblems
- Closed-form solutions (thus differentiable) for both subproblems
  - **Input:** initial depth, normal, confidence maps of depth and normal
  - **Output:** refined depth and normal

Validating the solver as a post-processing module

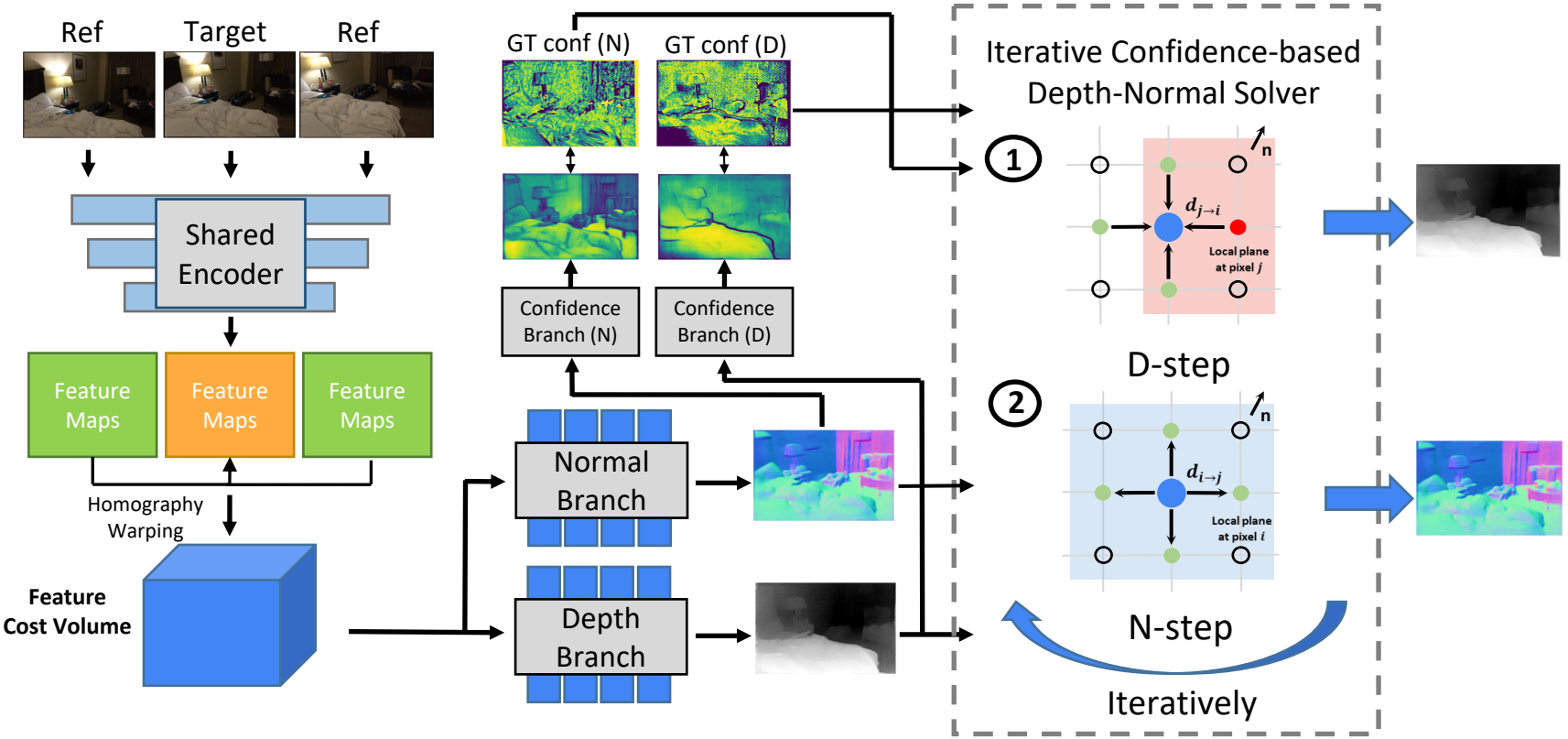
Image

Input depth

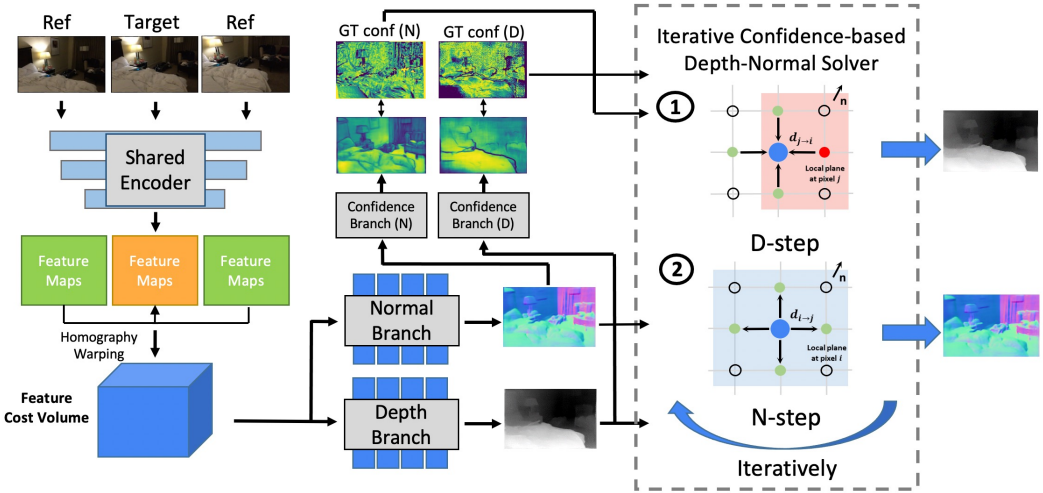
Output depth

- COLMAP MVS is used to obtain the initial depth and surface normal.
- Higher confidence is assigned to pixels with valid depth and normal projections at fusion.
- Our solver effectively propagates the reliable geometry into missing parts (e.g. low-texture regions) and largely improves reconstruction completeness.

Deep Multi-view Stereo System

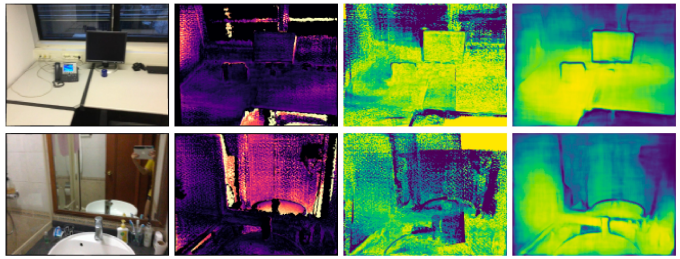


Deep Multi-view Stereo System

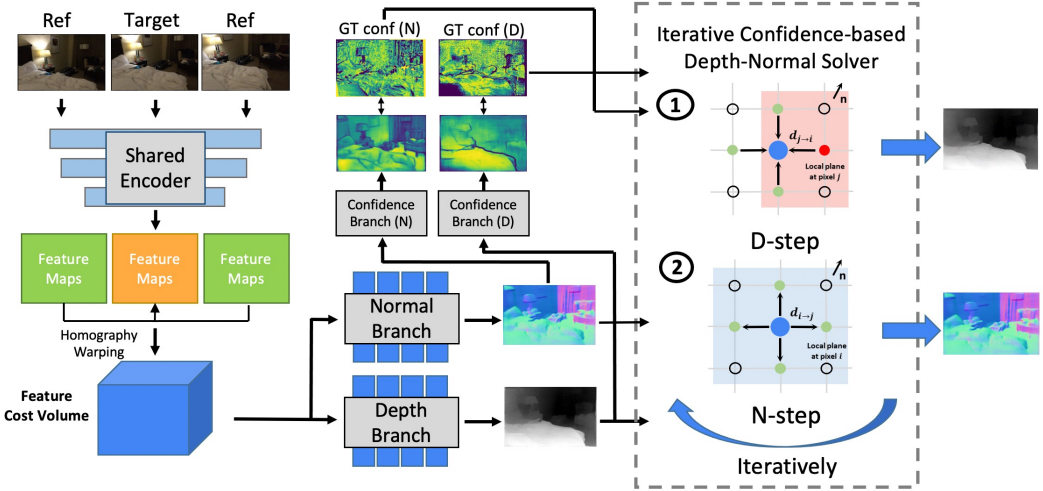


- Cost volume based initial depth and surface normal prediction
- Learning confidence prediction of the initial geometry
- Joint training by integrating the differentiable iterative solver

Deep Multi-view Stereo System



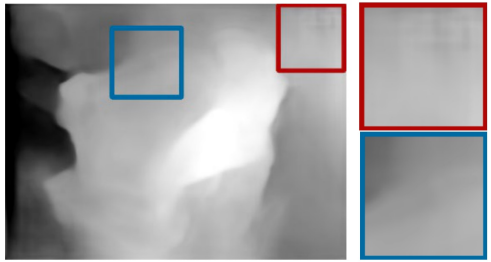
(a) Image (b) Depth Error (c) Geometric Conf (d) Deep Conf



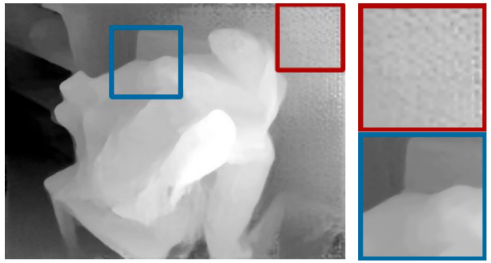
- Groundtruth confidence at training
  - GT confidence is calculated by comparing the predictions with groundtruth depth/normal.
- Hybrid confidence at inference
  - Deep confidence + Geometric confidence.
  - Deep confidence is obtained by confidence branches.
  - Geometric confidence is obtained by depth reprojection check.

Deep Multi-view Stereo System

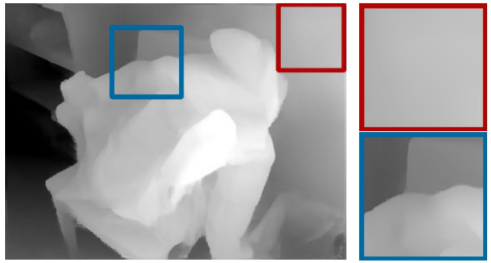
(a) Input RGB image



(b) w/o. solver



(c) w. solver only at training



(d) w. solver at training and inference

- **w/o. solver:**
  rough geometry on both regions
- **w. solver only at training:**
  accurate in textured region, “noisy” in texture-less region
- **w. solver at training and inference:**
  fine geometry on both regions

Results - MVS System - Depth

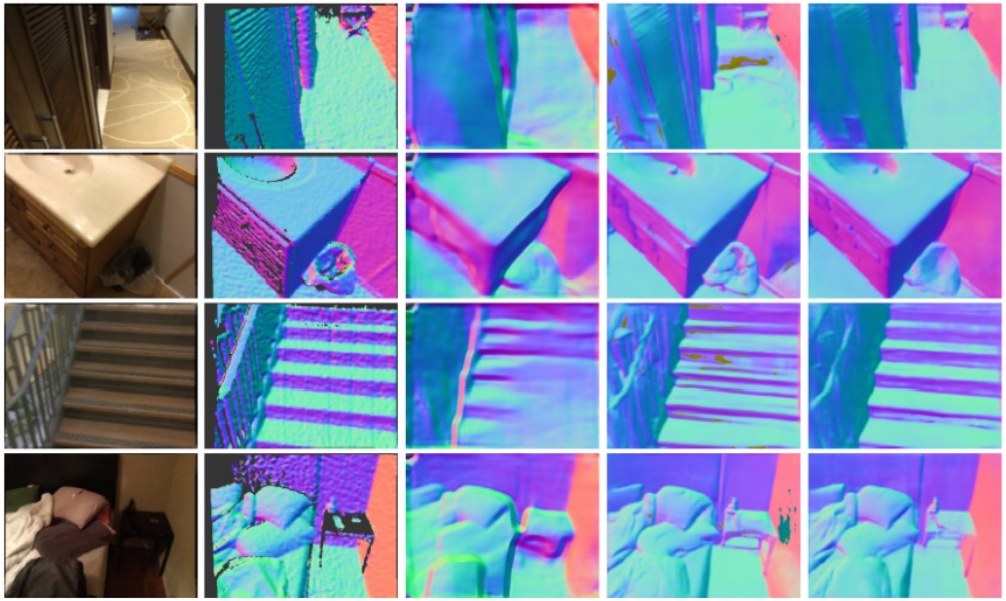
Method	Abs Rel	Abs Diff	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
MVDepth [46]	0.1053	0.1987	0.0634	0.3026	0.1490	0.8817	0.9723	0.9924
MVDepth (FT)	0.1014	0.1891	0.0476	0.2850	0.1390	0.8930	0.9764	0.9941
GP-MVS [20]	0.0920	0.2283	0.0644	0.4436	0.1560	0.8918	0.9629	0.9918
GP-MVS (FT)	0.0787	0.2008	0.0518	0.4009	0.1394	0.9134	0.9643	0.9931
NeuralRGBD [31]	0.0871	0.1710	0.0409	0.2693	0.1324	0.9150	0.9785	0.9925
CNM [32]	0.1119	0.2101	0.0510	0.2970	0.1485	0.8686	0.9724	0.9930
DPSNet [22]	0.1164	0.1992	0.0606	0.3065	0.1602	0.8569	0.9575	0.9884
DPSNet (FT)	0.0910	0.1807	0.0410	0.2697	0.1291	0.9008	0.9787	0.9952
NAS [28]	0.0795	0.1597	0.0323	0.2357	0.1112	0.9284	0.9862	0.9966
DELTAS [42]	0.0738	0.1380	0.0245	0.2051	0.1021	0.9473	0.9890	0.9976
Ours	**0.0665**	**0.1281**	**0.0240**	**0.1995**	**0.0990**	**0.9489**	**0.9896**	**0.9978**

ScanNet dataset

Method	Abs Rel	Abs Diff	Sq Rel	RMSE	$\delta < 1.25$
MVDepth [46]	0.0885	0.1467	0.0314	0.2313	0.9184
GP-MVS [20]	0.1087	0.1514	0.0827	0.2873	0.9170
N-RGBD [31]	0.0995	0.1530	0.0352	0.2361	0.9233
CNM [32]	0.1350	0.1873	0.0484	0.2619	0.8667
DPSNet [22]	0.0771	0.1290	0.0234	0.2045	0.9401
NAS [28]	0.0732	0.1241	0.0198	0.1893	0.9576
DELTAS [42]	0.1065	0.1528	0.0299	0.2138	0.9156
Ours	**0.0698**	**0.1130**	**0.0194**	**0.1770**	**0.9681**

RGB-D Scenes V2 dataset

Results - MVS System – Surface Normal



Method	Mean	Median	11.25°	22.5°	30°
CNM [32]	27.92	22.12	27.43	52.16	63.44
NAS [28]	24.12	18.02	31.59	60.20	69.45
Ours	**22.30**	**16.75**	**34.80**	**64.39**	**75.11**

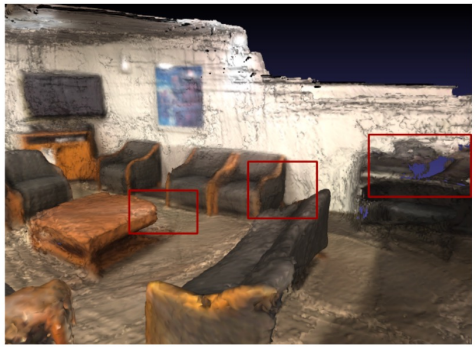
Table 2. Quantitative comparisons of surface normal estimation between our method and state-of-the-art methods [32, 28].

- We test our final surface normals on ScanNet dataset and achieves state-of-the-art performance on the task of surface normal estimation.

Results - MVS System - Reconstruction

(a) Groundtruth

(b) Ours



(c) DELTAS [43]

(d) DPS [23]

- Direct TSDF fusion results using predicted depths on ScanNet scenes.
- Our method reconstructs finer geometry details with less outliers.

2021 ICCV OCTOBER 11-17 VIRTUAL

A Confidence-based Iterative Solver of Depths and Surface Normals for Deep Multi-view Stereo

Wang Zhao*, Shaohui Liu*, Yi Wei, Hengkai Guo, Yong-Jin Liu