Learning Non-volumetric Depth Fusion using Successive Reprojections

Simon Donné, Andreas Geiger
Max-Planck-Institute for Intelligent Systems -- Autonomous Vision Group

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

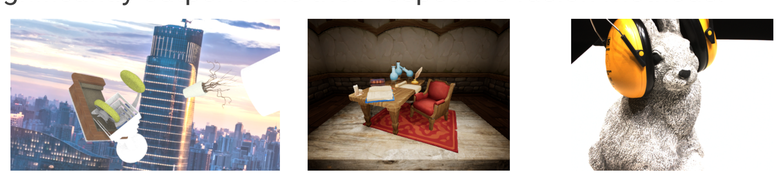
MAX-PLANCK-GESELLSCHAFT

Introduction

We tackle the problem of **depth map refinement and fusion** in the image domain. Whereas established techniques use 3D representations to perform depth fusion, an **image-based** representation lends itself better to learning-based approaches.

In **DeFuSR**, we reproject information from **neighbouring views** into the current view, and use this to refine the initial estimate. **Iterating** on this further improves the estimated depth. Additionally, we learn to estimate prediction **confidence**; the set of high-confidence points is a high-quality **fused point cloud**.

Our method was trained for both traditional (COLMAP) and learning-based (MVSNet) multi-view stereo approaches and significantly outperforms their respective fusion methods.

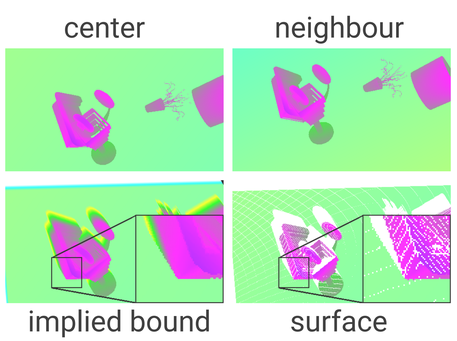


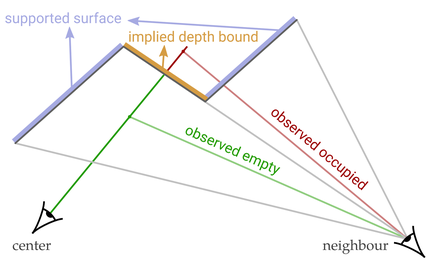
FlyingThingsMVS unrealDTU DTU

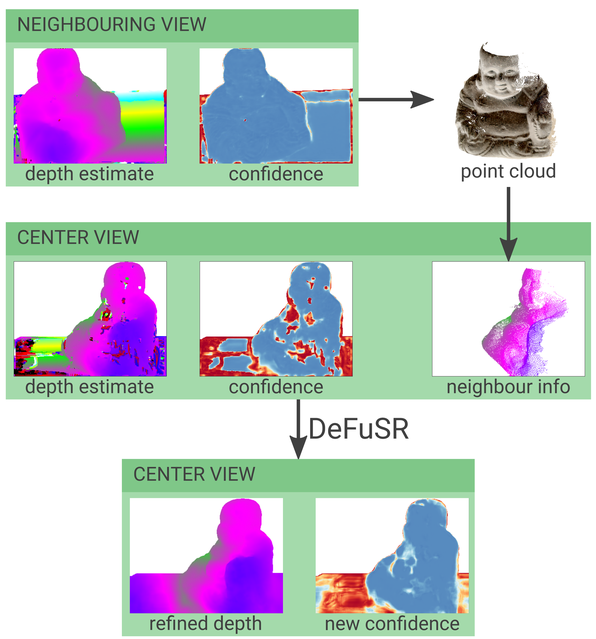
We pretrain on two synthetic datasets, FlyingThingsMVS and unrealDTU, and finetune on the DTU dataset. Furthermore, the depth refinement generalizes to **real scenes without retraining**.

Neighbour information

To encode the information of neighbouring views, we encode two types of information: **implied depth bounds** and **evidence-supported surfaces** (and their colour information).







Choice of neighbouring views

As expected, the number and choice of neighbours plays a significant role. More neighbours are better, but in practice limited by memory and computational complexity.

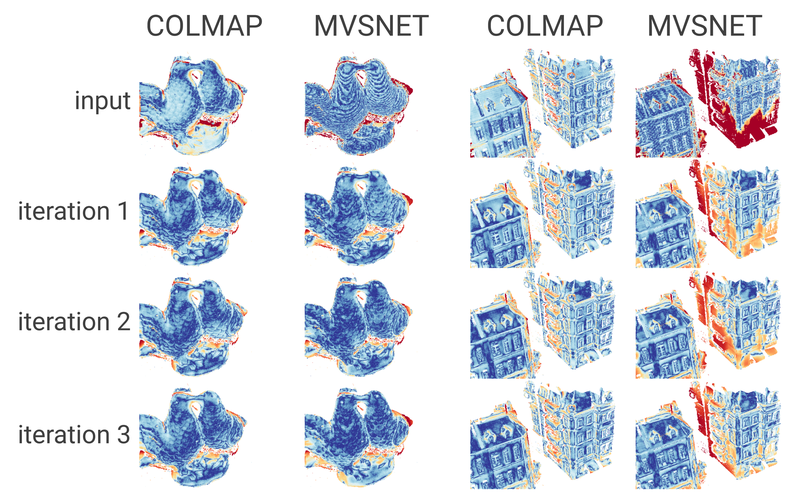
			neighbours		
		COLMAP	0	4	12
per-view	acc. (%)	66	91	92	89
	comp. (%)	40	28	31	45
	mean (%)	52	59	62	**67**
full	acc. (%)	73	81	84	80
	comp. (%)	72	66	64	84
	mean (%)	72	74	74	**82**

When it comes to neigbhour selection, we show that a mix of close-by and far-away neighbours is the best choice:

		COLMAP	nearest	mixed	furthest
per-view	acc. (%)	66	91	89	86
	comp. (%)	40	38	45	37
	mean (%)	52	64	**67**	62
full	acc. (%)	73	83	80	74
	comp. (%)	72	72	84	76
	mean (%)	72	78	**82**	75

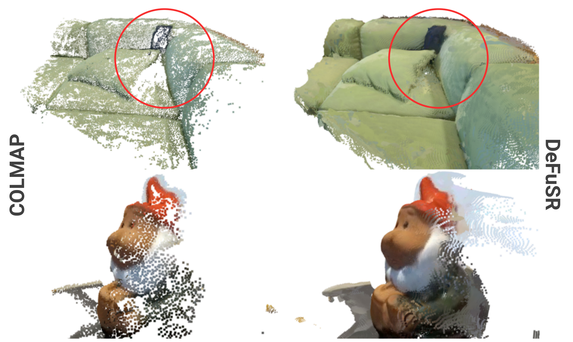
Number of iterations

Performing multiple iterations of our approach successively improves results, saturing quickly (we perform three iterations).



Generalization to real data

Our depth refinement generalizes to real data without retraining, as the same MVS technique (COLMAP) is used as an intermediate step.



simon.donne@tue.mpg.de

https://avg.is.tuebingen.mpg.de/research_projects/defusr --- Code and data available!