DELTAS: Depth Estimation by Learning Triangulation And densication of Sparse points

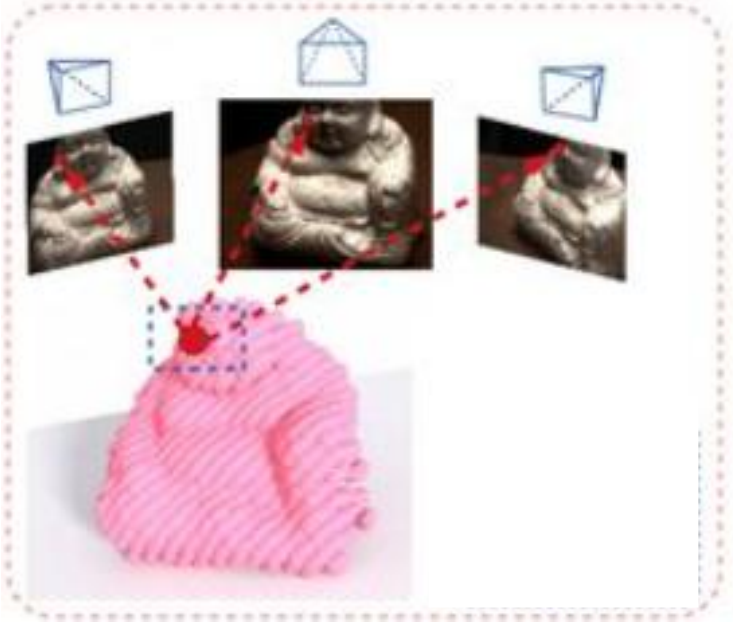
ECCV 2020 Magic leap

Introduction

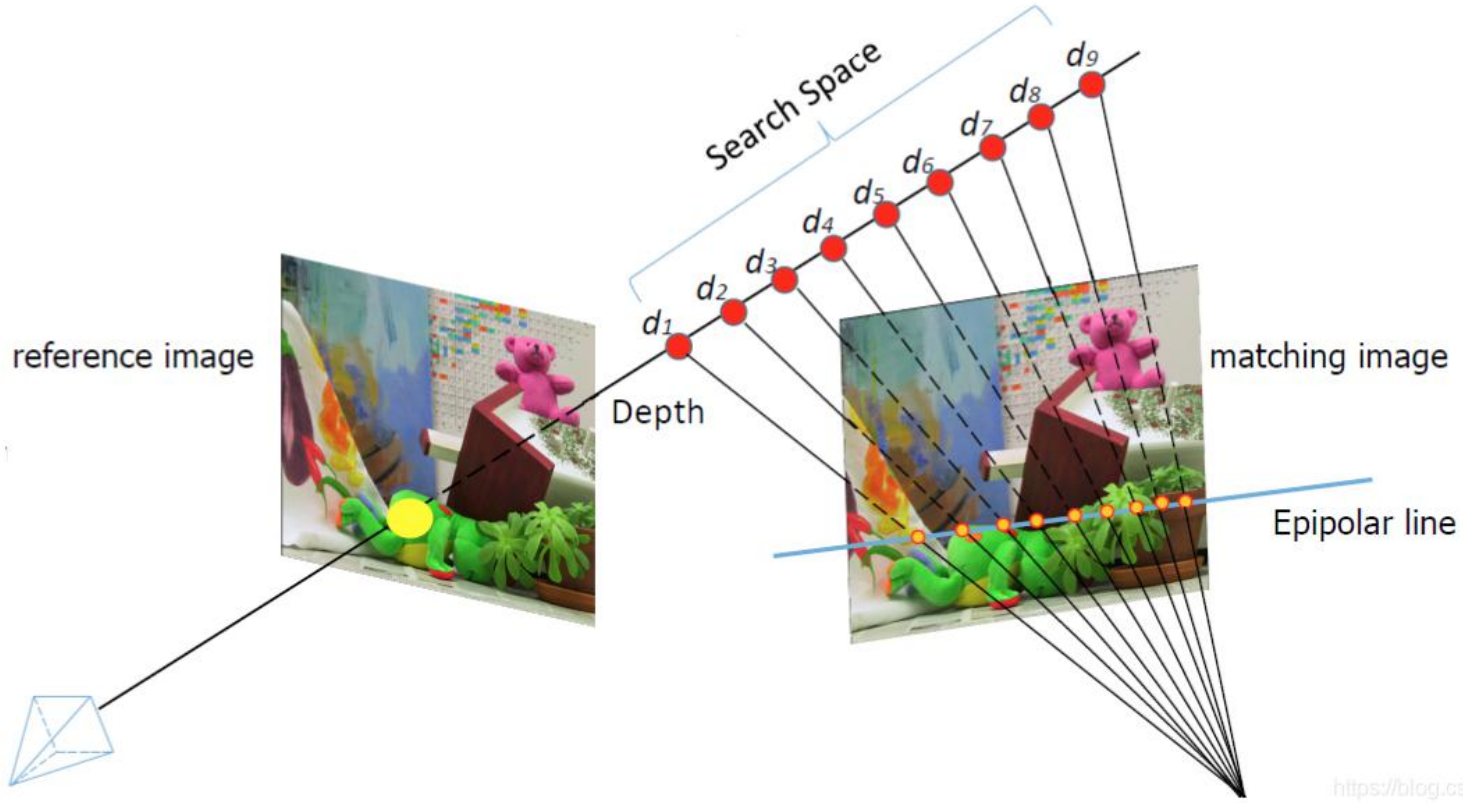
Problem setting: multi-view stereo

Input: Images/Videos; Camera intrinsic & extrinsic

Output: the depth map of every image / 3D points cloud



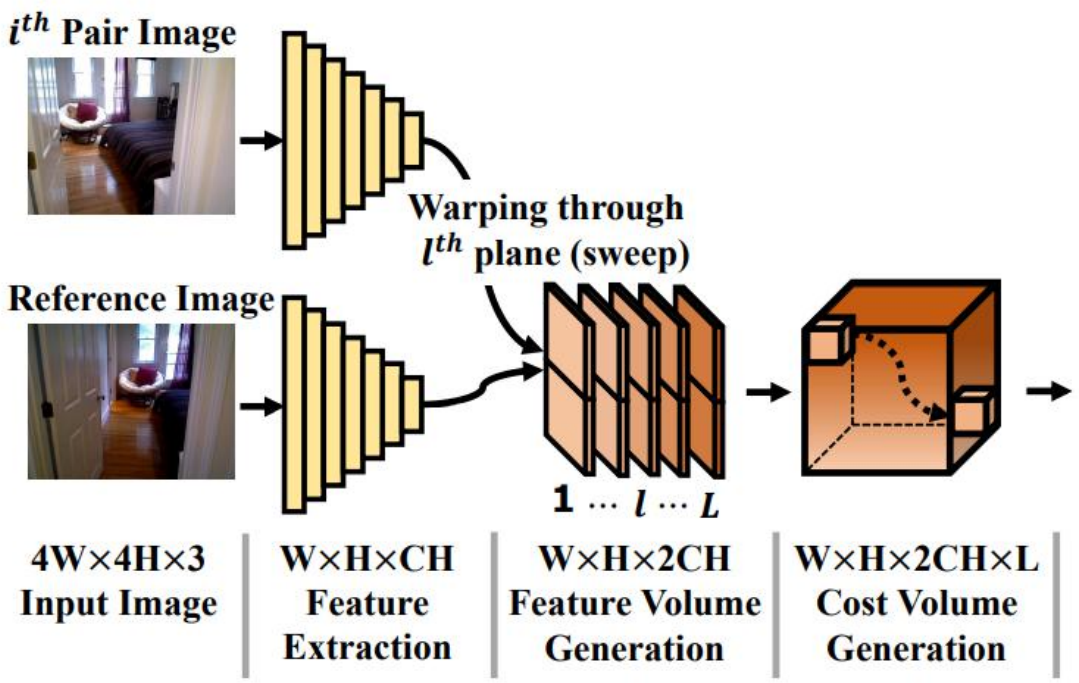
Prerequisite: Epipolar Search for MVS

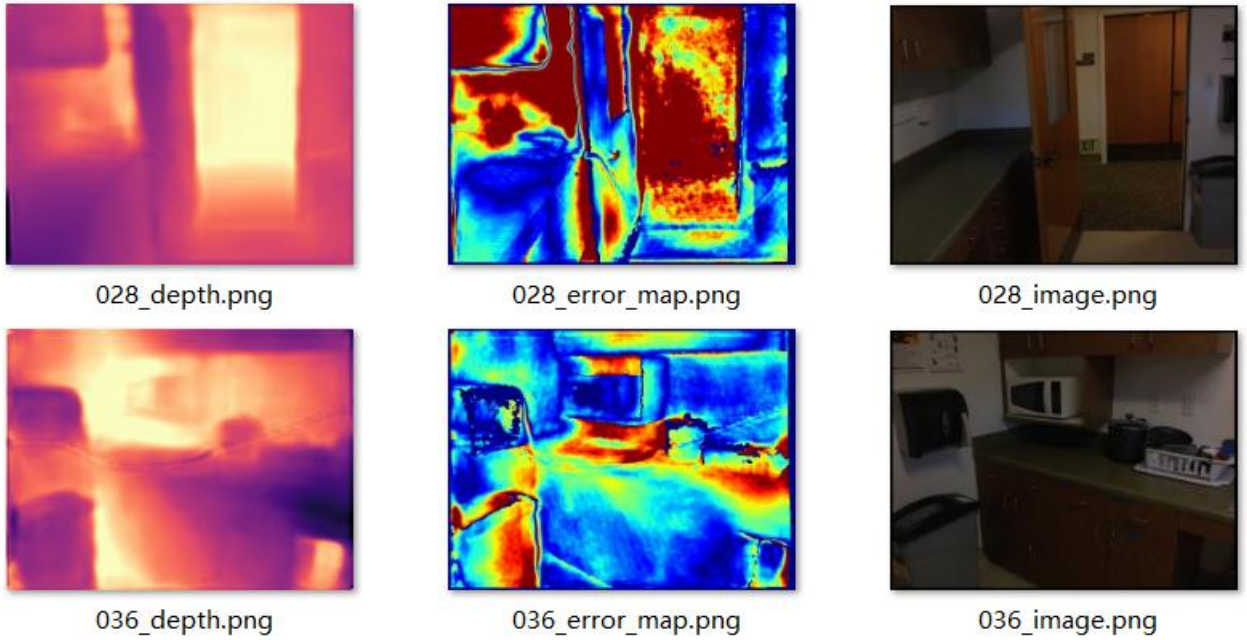


$$\begin{cases} x = f\frac{X}{Z} \\ y = f\frac{Y}{Z} \end{cases}$$

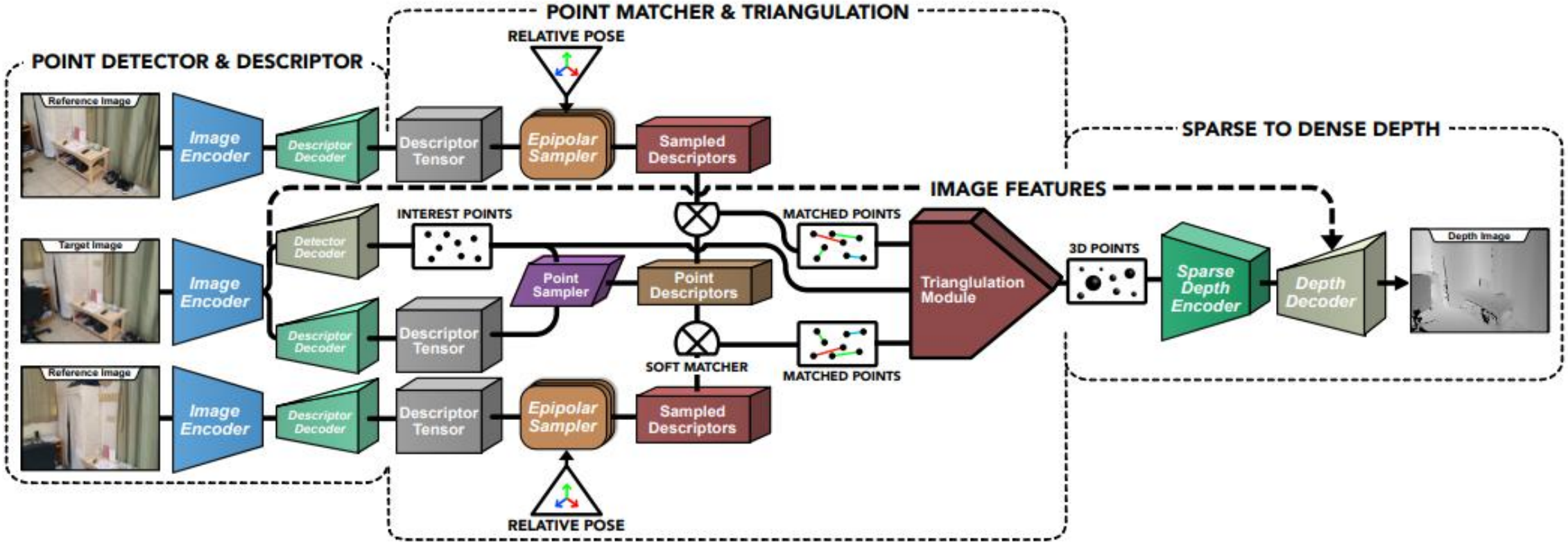
Motivation

Deep MVS Method highly rely on 3D cost volume, which is time & memory consuming.
Revisited the traditional geometry method!

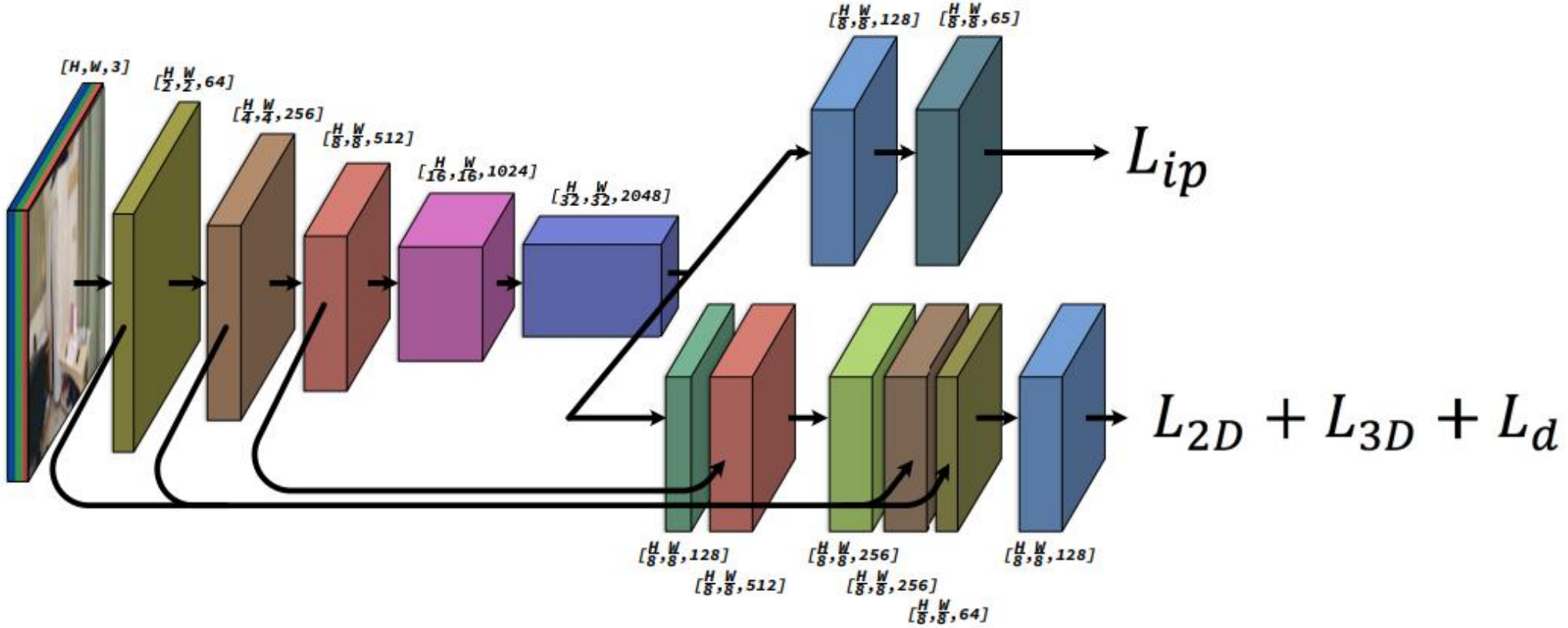




Method: Pipeline



Method: Interest point detector and descriptor



- Train the interest point detector network by **distilling the output of the original SuperPoint network** and the descriptors are trained by **the matching formulation**.

Method: Matching and Triangulation

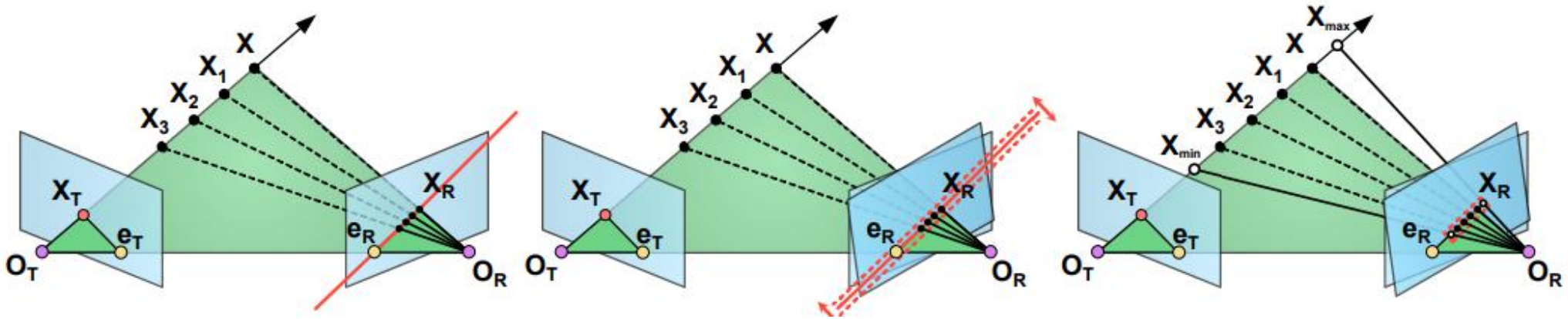
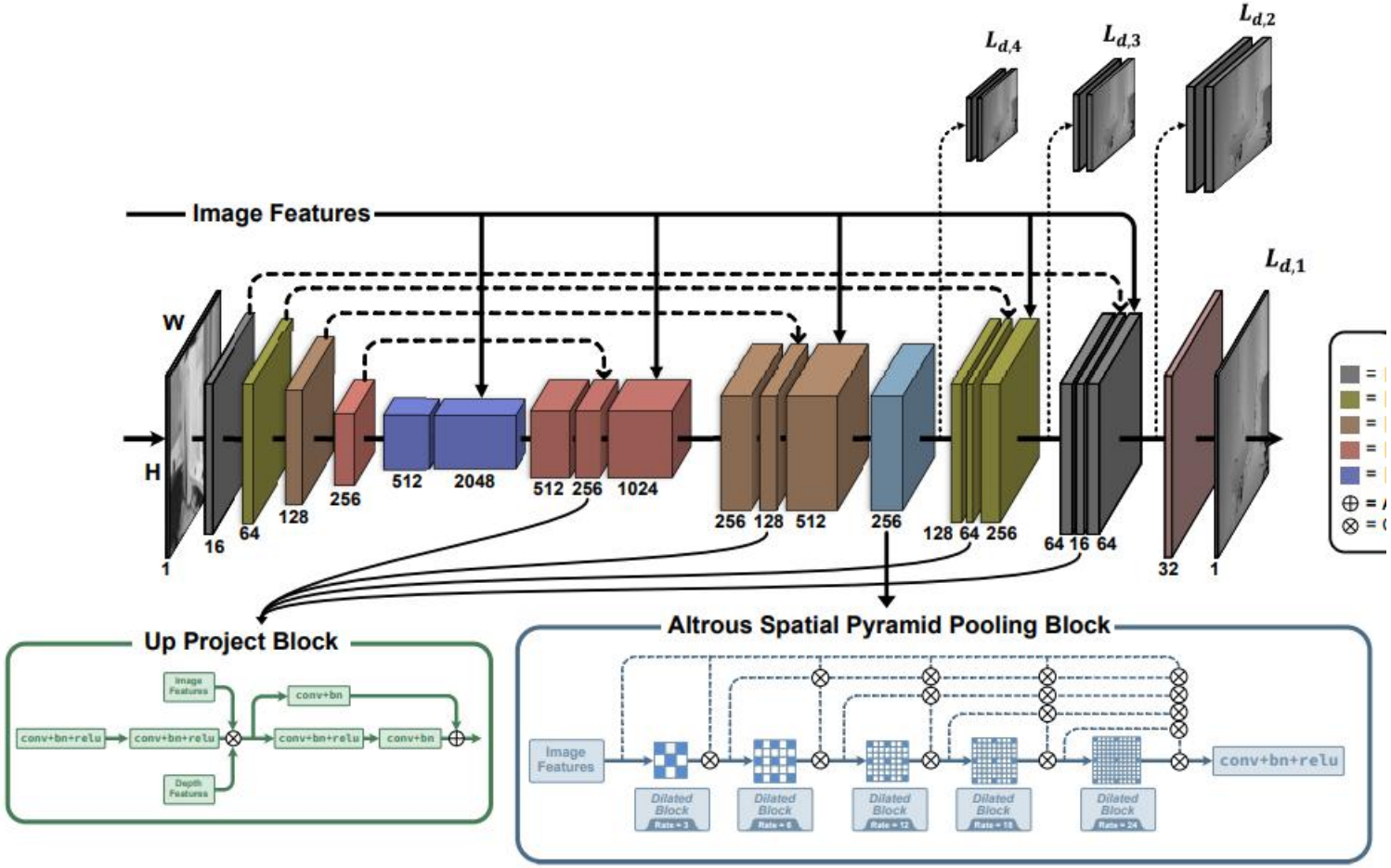


Fig. 3. Left: Epipolar sampling; Middle: Offset sampling due to relative pose error; Right: Constrained depth range sampling

1. Matching is from argmax operation which is differentiable

2. Triangulation: Key idea is to minimize the reprojection error

Method: Densification of sparse point



Experiments

ScanNet

	Abs Rel	Abs	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
GPMVS	0.1306	0.2600	0.0944	0.3451	0.1881	0.8481	0.9462	0.9753
GPMVS-FT	0.1079	0.2255	0.0960	0.4659	0.1998	0.8905	0.9591	0.9789
MVDepth	0.1191	0.2096	0.0910	0.3048	0.1597	0.8690	0.9599	0.9851
MVDepth-FT	0.1054	0.1911	0.0970	0.3053	0.1553	0.8952	0.9707	0.9895
DPS	0.1470	0.2248	0.1035	0.3468	0.1952	0.8486	0.9474	0.9761
DPS-FT	0.1025	0.1675	0.0574	0.2679	0.1531	0.9102	0.9708	0.9872
Ours	**0.0932**	**0.1540**	**0.0506**	**0.2505**	**0.1426**	**0.9287**	**0.9767**	**0.9893**

ScanNet for different views

Method	**2 Frames**			**4 Frames**			**5 Frames**			**7 Frames**		
	AbR	Abs	SqR	AbR	Abs	SqR	AbR	Abs	SqR	AbR	Abs	SqR
GPN	0.112	0.233	0.101	0.109	0.226	0.100	0.107	0.226	0.112	0.109	0.230	0.116
MVN	0.126	0.238	0.471	0.105	0.191	0.078	0.106	0.192	0.071	0.108	0.195	0.067
DPS	**0.099**	0.181	0.062	0.102	0.168	0.057	0.102	0.168	0.057	0.102	0.167	0.057
Ours	0.106	**0.173**	**0.057**	**0.090**	**0.150**	**0.049**	**0.088**	**0.147**	**0.048**	**0.087**	**0.144**	**0.043**

Experiments

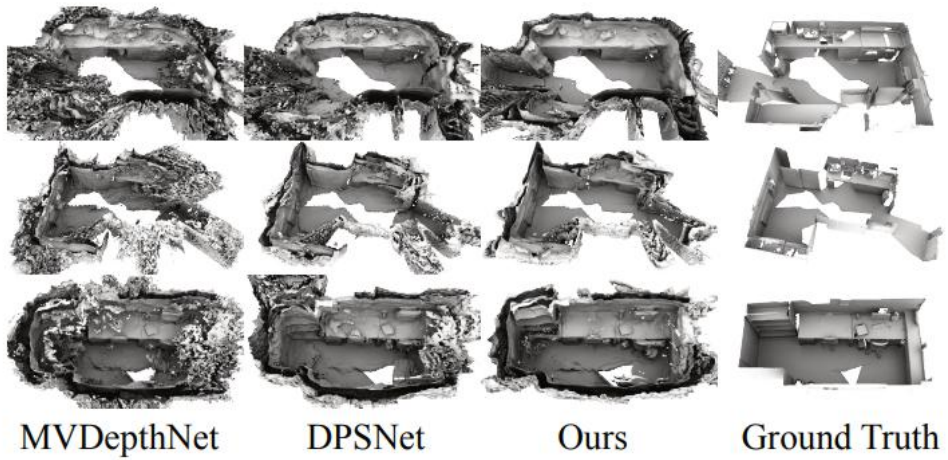


Fig. 5. 3D scene reconstruction using predicted depth over the full sequence.

Table 4. Performance of depth estimation on Sun3D. We use sequences of length 2.

	Abs Rel	Abs	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
MVDepth	0.1377	0.3199	0.1564	0.4523	0.1853	0.8245	0.9601	0.9851
MVDepth-FT	0.3092	0.7209	4.4899	1.718	0.319	0.7873	0.9117	0.9387
DPS	0.1590	0.3341	0.1564	0.4516	0.1958	0.8087	0.9363	0.9787
DPS-FT	0.1274	0.2858	0.0855	0.3815	0.1768	0.8396	0.9459	0.9866
Ours	**0.1245**	**0.2662**	**0.0741**	**0.3602**	**0.1666**	**0.8551**	**0.9728**	**0.9902**