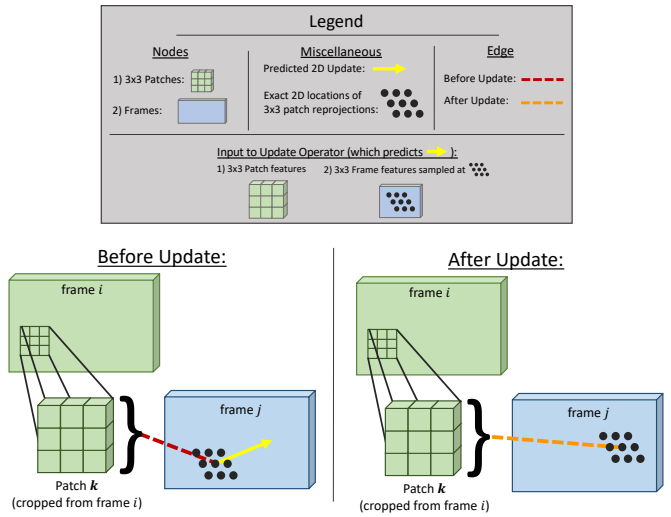

Figure L: The patch graph, focusing on a single pair of nodes (a patch and a frame), and the single edge connecting them. Note that the patch reprojections are computed using the poses from frames i and j, and the depth of patch $\mathbf{k}$

	Runtime	Uses Global Optimization?	MH01	MH02	MH03	MH04	MH05	V101	V102	V103	V201	V202	V203
D3VO	not reported	**NO**	train	train	**0.08**	train	0.09	train	train	0.11	-	0.05	0.19
ORB-SLAM3	20 FPS	YES	**0.016**	**0.027**	0.028	**0.138**	**0.072**	**0.033**	**0.015**	**0.033**	**0.023**	**0.029**	FAIL
Ours (Default)	60FPS	**NO**	0.087	0.055	0.158	0.137	0.114	0.050	0.140	0.086	0.057	0.049	0.211
Ours (Fast)	**120FPS**	**NO**	0.101	0.067	0.177	0.181	0.123	0.053	0.158	0.095	0.095	0.063	0.310

Table A: Results on EuRoC. As expected, methods with global optimization (ORB-SLAM3), when it doesn't fail catastrophically, outperform VO methods on EuRoC, which has many loops. Note that ORB-SLAM3 fails catastrophically on one sequence where ours fails on zero. D3VO performs very well, which is expected as it performs offline-pretraining on the evaluation scenes, meaning that it effectively performs offline 3D reconstruction in advance. In contrast, our approach is evaluated on the EuroC dataset zero-shot.

	fr1/desk	fr2/xyz	fr3/office	Requires RGB-D?	Runtime	Uses Global Optimization?	Un-Bounded memory?
iMAP	4.9	2.0	5.8	YES	10-FPS	YES	YES
NICE-SLAM	2.8	2.4	3.0	YES	21-FPS	PARTIALLY	YES
Li et al.	**2.0**	0.6	**2.3**	**NO**	14-FPS	PARTIALLY	YES
Ours (Default)	2.53	**0.47**	7.08	**NO**	**60-FPS**	**NO**	**NO**
Ours (Fast)	2.62	0.54	8.31	**NO**	**120-FPS**	**NO**	**NO**

Table B: Error on TUM-RGBD. Results are in RMSE ATE [cm], lower is better. All methods except ours use global optimization, as opposed to DPVO which is a strict VO system. DPVO will therefore never run out of memory even on arbitrary-length videos of unboudned scenes. iMAP and NICE-SLAM also require RGB-D sequence, whereas Li et al. and ours expect only an RGB sequence. Our approach runs 3x-8x faster.

	V101	V102	V103	V201	V202	V203	Runtime	Uses Global Optimization?	Un-bounded memory?
Li et al.	0.068	**0.079**	FAIL	**0.053**	0.178	FAIL	14-FPS	PARTIALLY	YES
Ours (Default)	**0.050**	0.140	**0.086**	0.057	**0.049**	**0.211**	60 FPS	**NO**	**NO**
Ours (Fast)	0.053	0.158	0.095	0.095	0.063	0.310	**120 FPS**	**NO**	**NO**

Table C: Error on EuRoC MAV. Results are in ATE [m], lower is better. Our approach outperforms Li et al. on 4/6 of the sequences, two of which result in a catastrophic failure for Li et al. Our approach is a strict VO system, so unlike Li et al. it uses no global optimziation, and will never run out of memory even on arbitrary-length videos of unbounded scenes.

Sequence	Xue et al.	Ours (default)	Ours (fast)
fr2/desk	**0.183**	0.349	0.347
fr2/pioneer_360	0.313	**0.099**	0.099
fr2/pioneer_slam	0.241	**0.089**	0.097
fr2/360_kidnap	**0.149**	0.332	0.549
fr3/cabinet	**0.193**	0.390	0.392
fr3/long_office_hou_valid	**0.017**	0.378	0.380
fr3/nostr_texture_near_loop	0.371	**0.370**	0.371
fr3/str_notexture_far	**0.046**	0.335	0.335
fr3/str_notexture_near	**0.069**	0.215	0.216
Trained on the test scenes?	YES	**NO**	**NO**
Runtime	not reported.	**60 FPS**	**120 FPS**

Table D: Error on TUM-RGBD. Results are in RMSE Relative Pose Error (RPE) [m/s]. DPVO outperforms Xue et al. on three of the sequences, while they outperform us on the remaining sequences. Xue et al. is trained on sequences that share the same scenes with the test sequences, whereas our approach is trained only on synthetic data and is tested on TUM-RGBD zero-shot.

Method	Average ATE [m]	Runtime
SuperGlue + SuperPoint + COLMAP	0.340	0.75 FPS
Ours (default)	**0.130**	**60 FPS**

Table E: Results on the TartanAir test set from the ECCV 2020 SLAM competition. DPVO achives 62% lower error than the approach based on SuperPoint+Superglue, and also runs 80x faster.

Sequence	DF-VO	Ours (default)	Ours (fast)
fr2/desk	**0.306**	0.349	0.347
fr2/pioneer_360	0.599	**0.099**	**0.099**
fr2/pioneer_slam	0.585	**0.090**	0.097
fr2/360_kidnap	0.745	**0.333**	0.549
fr3/cabinet	0.447	**0.390**	0.392
fr3/long_office_hou_valid	**0.227**	0.378	0.380
fr3/nostr_texture_near_loop	0.564	**0.371**	**0.371**
fr3/str_notexture_far	0.505	**0.335**	**0.335**
fr3/str_notexture_near	0.603	**0.215**	0.216
Avg	0.509	**0.285**	0.310
Runtime	not reported	60 FPS	**120 FPS**

Table F: Results on TUM-RGBD, measured using RMSE Relative pose error (RPE) [m/s]. Our approach outperforms DF-VO.

	MH000	MH001	MH002	MH003	MH004	MH005	MH006	MH007
ORB-SLAM1	1.30	**0.04**	2.37	2.45	FAIL	FAIL	21.47	2.73
ORB-SLAM3	15.44	2.92	13.51	8.18	2.59	21.91	11.70	25.88
Ours (Default)	**0.21**	**0.04**	**0.04**	0.08	**0.58**	**0.17**	0.11	0.15
Ours (Fast)	0.34	0.05	0.06	**0.07**	0.81	0.41	**0.09**	**0.14**

Table G: Results on the TartanAir test set, measured in ATE [m]. Our method outperforms both ORB-SLAM 1 and 3.