

Improving SfM for Compact Multi-Camera Robots in Low-Light Environments

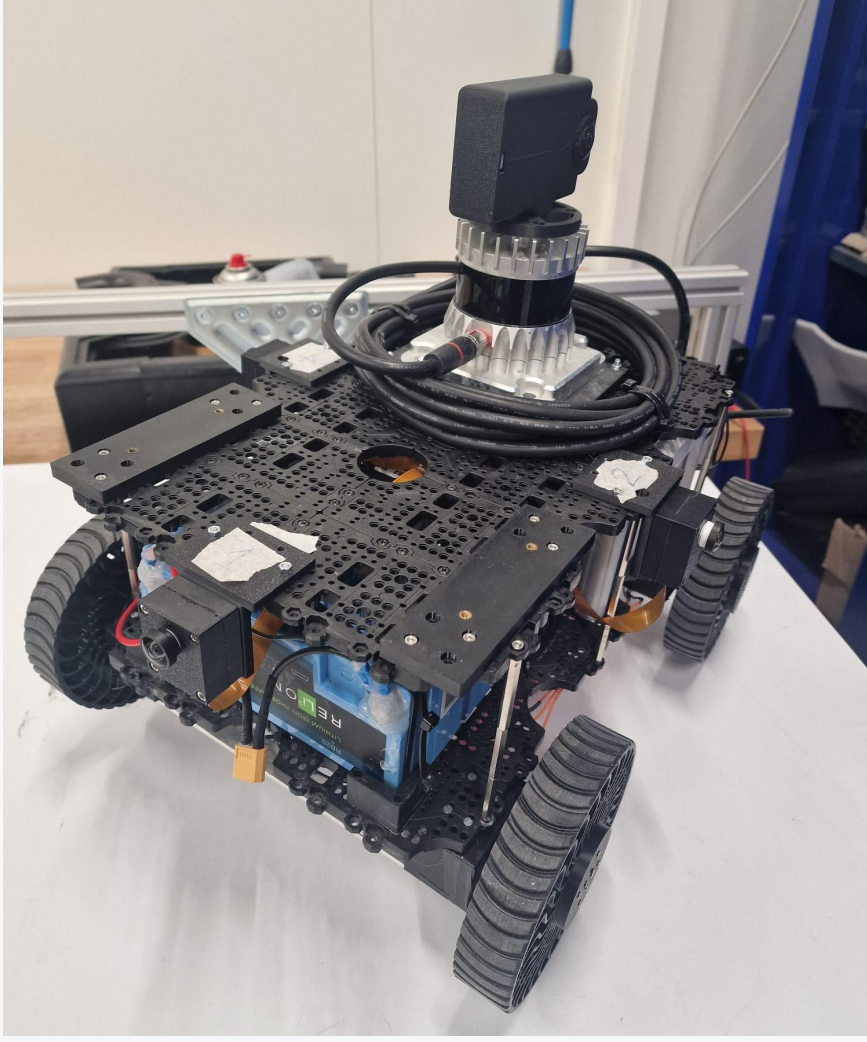
A CLAHE-centred ablation that unlocks full-trajectory reconstruction inside an unlit box-girder bridge

Stefano Mutti Diego Gitardi Simone Sabbadini Anna Valente

SUPSI – University of Applied Sciences and Arts of Southern Switzerland, stefano.mutti@supsi.ch

ICRA 2026 Workshop on Field Robotics

Platform & Mission



Compact tracked inspection robot, SUPSI.

Compact mobile robot deployed inside box-girder bridge cavities for structural inspection.

- **Perception:** 4 global-shutter cameras for 360° coverage, co-moving LED illumination, 9-DOF IMU for future pose priors.
- **Mobility:** tracked drive running on 3D-printed TPU wheels — high-friction, shock-absorbing compliance for the rough interior terrain (weld seams, bolt heads, debris, damp steel).
- **Envelope:** fits the typical 60–90 cm access hatch of a European box-girder bridge.
- **Mission:** a dense photometric 3D reconstruction of the hollow shaft for crack and corrosion survey.

SfM pose recovery is the *prerequisite* step for downstream 3D Gaussian Splatting — solve pose in the dark first, photometry second.

Root Cause

- **Intra-frame non-uniformity.** LED spot $\Rightarrow$ bright centre, dark edges; SIFT peak threshold starves the periphery.
- **Inter-frame variation.** Light rigidly attached to the robot — the same surface patch is lit very differently in two consecutive frames.
- **Descriptor ambiguity.** Low-contrast patches collapse to a common gradient pattern; matching breaks.

The failure is photometric, not geometric, so the fix belongs upstream of matching.

Bridge Dataset



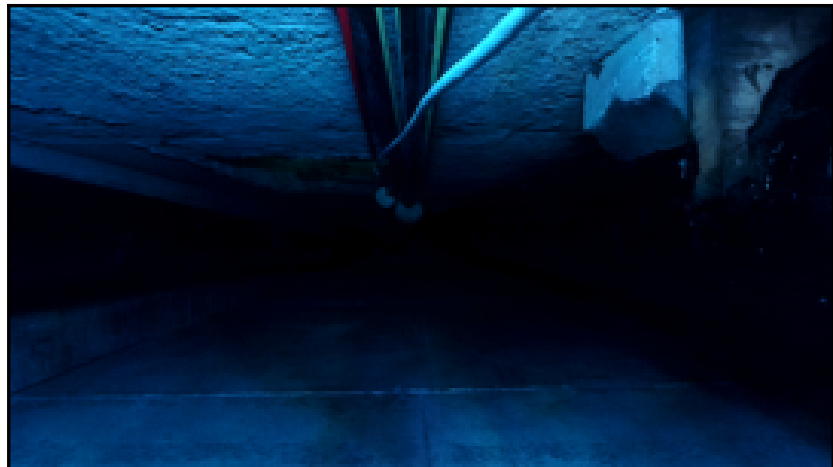
Bridge exterior.



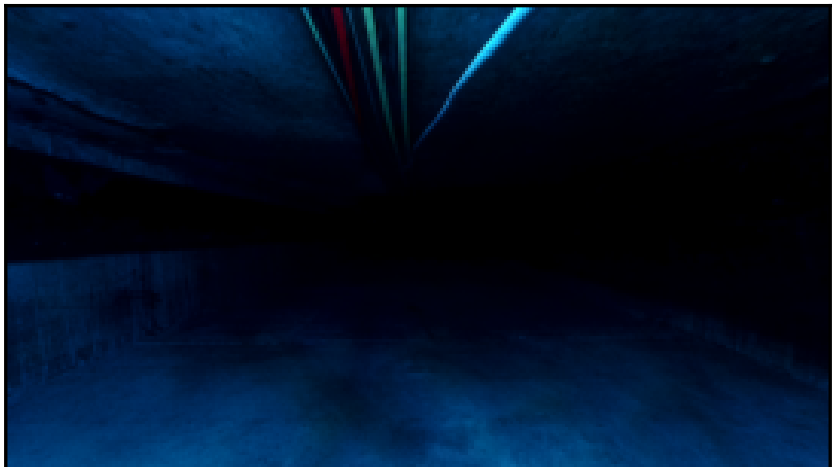
Interior — person for scale.

Bridge box-girder interior: 4 samples of 298 frames

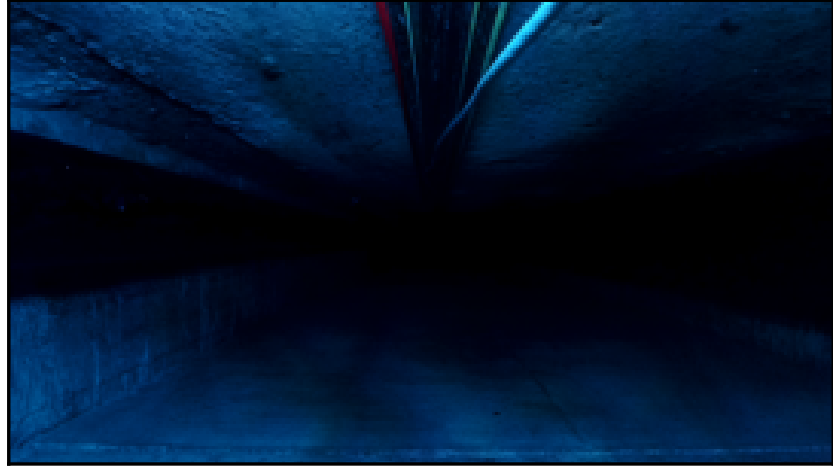
Frame 1/298



Frame 100/298



Frame 199/298



Frame 298/298



298 frames, 960×540, robot traverses the box-girder interior: zero ambient light, confined geometry, textureless concrete and steel. Sequential matching (overlap = 15); 5-config ablation, vanilla COLMAP [3] vs. our patched colmap_new.

Why This Matters

- **Co-moving-light regime** is shared by bridges, tunnels, caves, mines, pipes.
- **Pose is the bottleneck for 3D-GS.** Gaussian Splatting, NeRF, and dense MVS all start from SfM poses.
- Fix sits **upstream of matching** $\Rightarrow$ plugs into COLMAP / hloc / learned-feature pipelines.

Ablation Pipeline

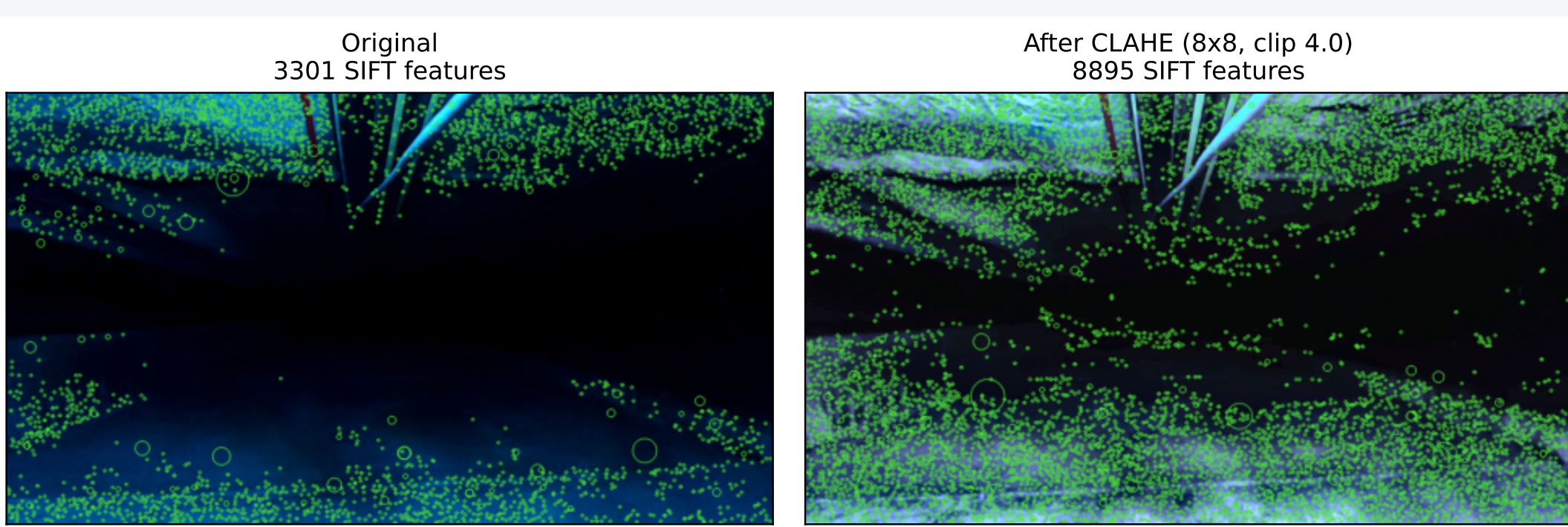
A Baseline	vanilla COLMAP 3.14 defaults, sequential matching overlap 15
1242 KP/img · 669 verified pairs · 28 imgs registered · 2962 3D points · 0.62 px reproj. err	
B + CLAHE preprocessing	luminance-channel adaptive hist. equalisation, 8×8 tiles, clip 4.0
1912 KP/img (+54%) · 754 pairs · 53 imgs · 5836 points · 0.37 px err (stochastic init — rerun also yielded 6 imgs)	
C + Adaptive SIFT threshold	$\tau_{\text{eff}} = \tau_0 \cdot \text{clamp}(\bar{l}/128, 0.1, 1.0)$, $\tau_0 = 0.004$
4260 KP/img (+243%) · 921 pairs · 54 imgs · 5659 points (dark frames back on budget)	
D + DSP-SIFT descriptors	domain-size pooling across 10 scales $\Rightarrow$ invariance to scale-varying illumination
11188 KP/img (+801%) · 1124 pairs · 116 imgs · 19071 points · 0.49 px err	
E + Relaxed matching / registration	ratio ρ 0.8 $\rightarrow$ 0.85, init inliers 100 $\rightarrow$ 50, abs.-pose 30 $\rightarrow$ 15, min inlier ratio 0.25 $\rightarrow$ 0.15
11188 KP/img · 2064 pairs (+208%) · 298 imgs (100% — full trajectory) · 54950 points · 0.55 px err	

Compound effect: 28 $\rightarrow$ 298 registered images (10.6 $\times$), +801% keypoints, 18.5 $\times$ denser 3D cloud (2962 $\rightarrow$ 54950 points) — each stage strictly improves on the previous; the full preset is the one that unlocks the full 298-frame trajectory.

(A) CLAHE Preprocessing

8×8 tile grid, clip limit 4.0 on the luminance channel. CLAHE [1] recovers gradient information from dark peripheral regions while preserving the well-lit centre. Also normalises *inter-frame* contrast: the same surface patch produces more consistent gradients across views.

What CLAHE Actually Does



Same bridge frame: **3301 SIFT features** without CLAHE (clustered at the bright centre) vs. **8895 features** after CLAHE (+170%), spread across the whole image.

(B) Adaptive SIFT Threshold

Scale the DoG peak threshold with per-image mean intensity $\bar{l}$:

$$\tau_{\text{eff}} = \tau_0 \cdot \text{clamp}\left(\frac{\bar{l}}{128}, 0.1, 1.0\right), \quad \tau_0 = 0.004.$$

Up to 10 $\times$ lower threshold on very dark frames; default recovered on well-exposed ones.

(C) DSP-SIFT Descriptors

DSP-SIFT [2] averages SIFT descriptors across multiple patch scales. Since the illumination pattern shifts between frames, fixed-scale contrast is inconsistent — pooling over scales makes descriptors more repeatable across views.

(D) Relaxed Matching

SIFT ratio test ρ : 0.8 $\rightarrow$ 0.85 · init. inliers 100 $\rightarrow$ 50 · abs.-pose inliers 30 $\rightarrow$ 15 · min inlier ratio 0.25 $\rightarrow$ 0.15. Reprojection filter kept at 4.0 px — geometric accuracy preserved, weaker-but-valid matches survive.

Limits & Next Steps

Limits. Single bridge run; no absolute-accuracy ground truth. Two-view initialisation is stochastic on feature-starved dark frames — the B-config rerun recovered 53 imgs, a first run had collapsed to 3.

Next. IMU-assisted pose initialisation (bypass the bimodal-init failure) · learned features (SuperPoint / LightGlue) as drop-in SIFT replacement under the same CLAHE front-end · multi-bridge validation with total-station ground truth · dense 3D-GS reconstruction fed by the E-preset poses.

Headline Numbers

28 $\rightarrow$ 298

registered images
(100% of the trajectory)

9 $\times$

keypoints / image
(1242 $\rightarrow$ 11188)

$\approx$ 0.55 px reprojection error

sub-pixel geometric accuracy preserved — ready for 3D-GS / dense MVS downstream

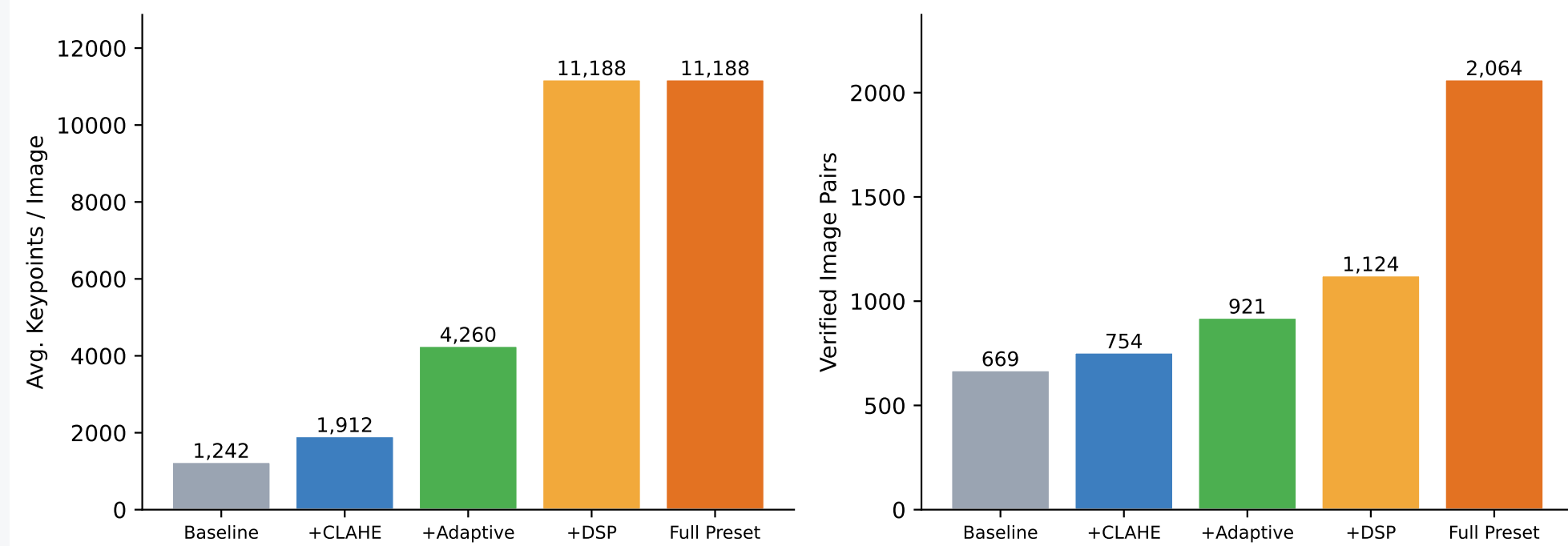
10.6 $\times$

reconstruction coverage
(vs. vanilla COLMAP)

18.5 $\times$

denser 3D cloud
(2962 $\rightarrow$ 54950 pts)

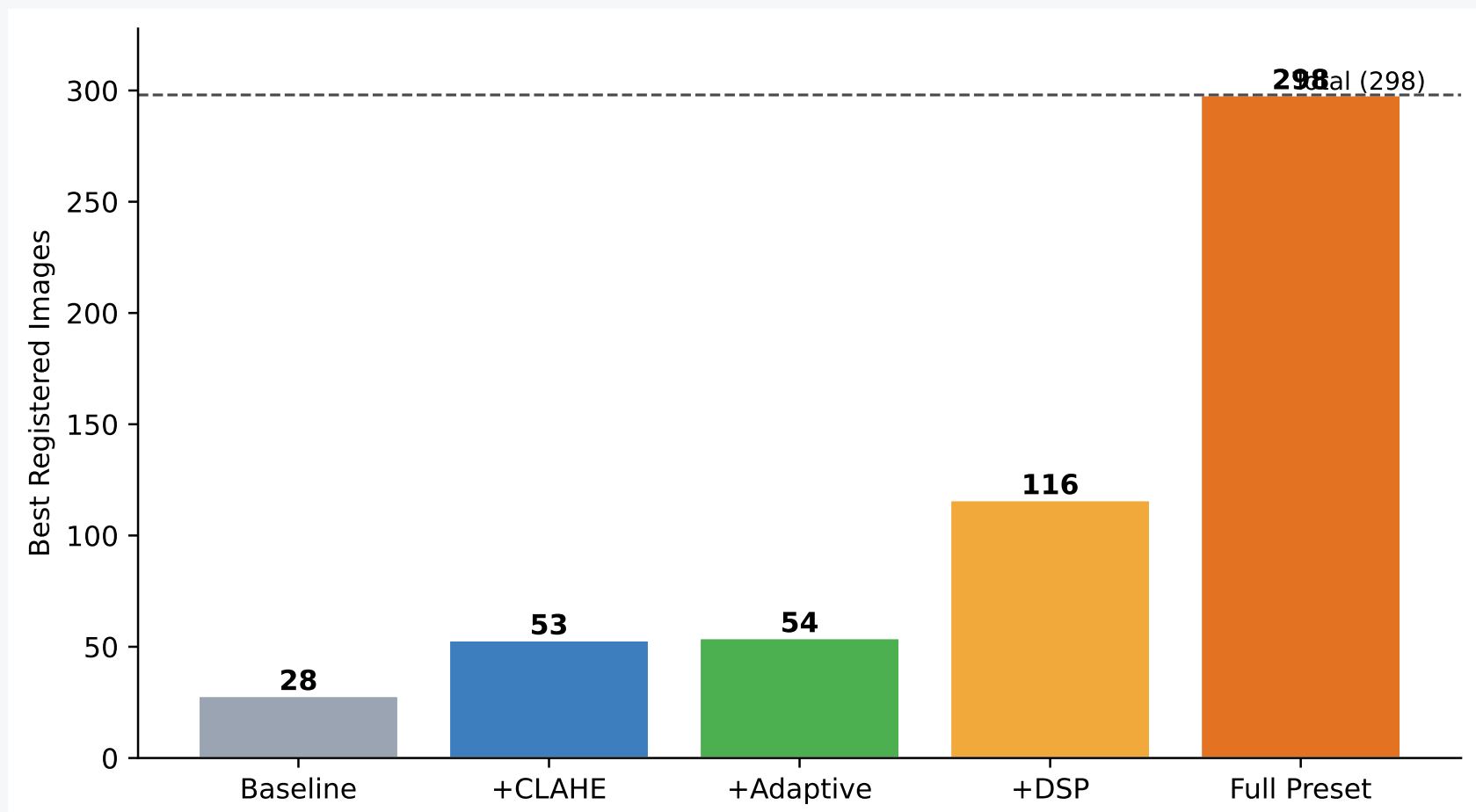
Features & Matching



Config	avg KP	med. KP	Pairs	Match %	Δ KP
A: Baseline	1242	1042	669	15.4	—
B: +CLAHE	1912	1755	754	17.3	+54%
C: +Adaptive	4260	4257	921	21.2	+243%
D: +DSP	11188	11210	1124	25.8	+801%
E: Full	11188	11210	2064	47.4	+801%

Match % = verified pairs / 4350 possible (overlap = 15 on 298 imgs).

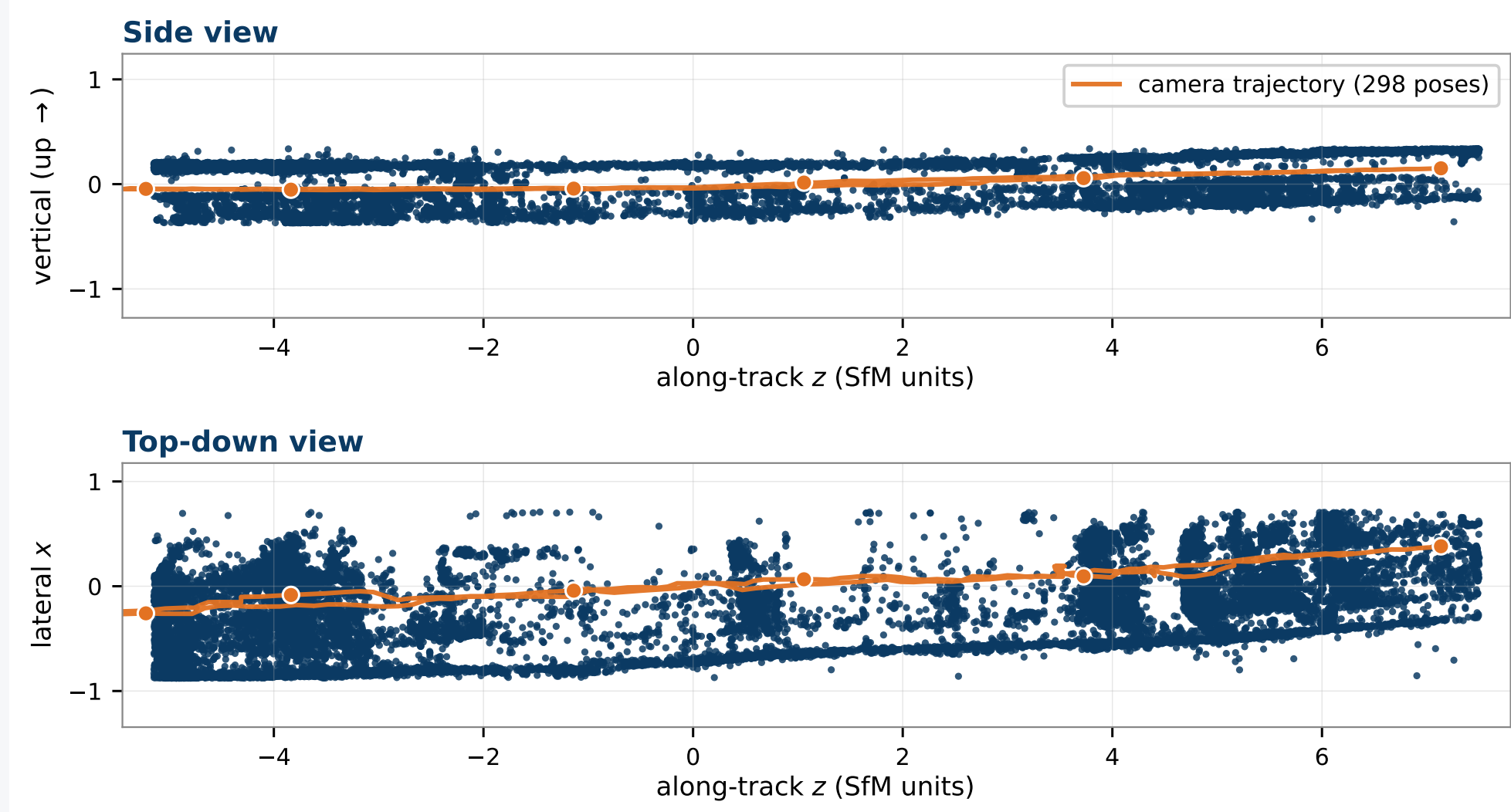
Reconstruction Coverage



Config	Reg.	%Reg	Points	Pts/img	Err (px)
A: Baseline	28	9.4	2962	106	0.62
B: +CLAHE	53	17.8	5836	110	0.37
C: +Adaptive	54	18.1	5659	105	0.48
D: +DSP	116	38.9	19071	164	0.49
E: Full	298	100.0	54950	184	0.55

%Reg = registered / 298; Pts/img shows per-image track density (higher = richer reconstruction).

Sparse Cloud & Poses



Output of run **E (Full Preset)**, sparse/1 submodel: **54950** 3D points and **298/298** registered camera poses along the box-girder axis — exactly the input a downstream 3D Gaussian Splatting or dense-MVS pipeline needs. No other config (A–D) recovered the full trajectory.

COLMAP Preset

Image pre-processing	
CLAHE tile / clip	8×8 / 4.0
Colour space	CIELAB, L-channel only
SIFT extraction	
Base peak threshold	$\tau_0 = 0.004$
Adaptive clamp	$\bar{l}/128 \in [0.1, 1.0]$
Descriptor	DSP-SIFT (10 scales)
Matching / mapping	
Matcher	sequential, overlap 15
SIFT ratio test	$\rho = 0.85$ (default 0.8)
Init. / abs.-pose inliers	50 / 15 (default 100 / 30)
Min inlier ratio	0.15 (default 0.25)
Reproj. error filter	4.0 px (unchanged)

Drop-in for any COLMAP incremental-SfM run on dark, robot-carried-illumination sequences. One binary flag group, no changes to geometry or bundle adjustment.

\$ colmap_new feature_extractor \



Workshop on Field Robotics, ICRA 2026

norlab-ulaval.github.io/icra_workshop_field_robotics

Contact: stefano.mutti@supsi.ch · SUPSI ISTePS

Refs:

- [1] Zuiderveld, CLAHE (1994)
- [2] Dong et al., DSP-SIFT (2015)
- [3] Schönberger & Frahm, COLMAP (2016)