ReFineR: Single image to dense point cloud registration at pixel level with point splatting

[1]Center for Robotics
60 Boulevard Saint-Michel
75006 Paris

Ugo Insalaco[1, 2]
ugo.insalaco@minesparis.psl.eu

Jean-Emmanuel Deschaud[1]
jean-emmanuel.deschaud@minesparis.psl.eu

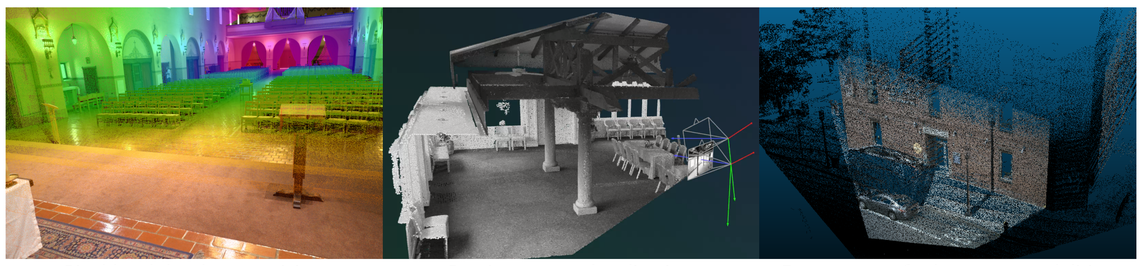
Pierre Dellenbach[2]
pdellenbach@chapsvision.com

[2]ChapsVision
4 Rue du Port aux Vins
92150 Suresnes

BMVC 2026

Context

On robotic platforms, residual reprojection errors of several pixels are commonly observed even after calibration due to temporal desynchronization, mechanical vibrations, or inaccuracies in point cloud reconstruction. Such errors can lead to decimeter-level geometric discrepancies at moderate distances and significantly affect multiple downstream tasks, such as dense cross-modal supervision, texture mapping, and semantic label transfer. Concretely, this work focuses on estimating up to pixel precision, the different camera parameters (localization, orientation, focal distance, distortion) from any kind of point cloud acquired (LiDAR, RGBD cameras...).



Pixel-level depth supervision | Precise camera localization | Feature transfer between modalities

Fig. 1: Example applications of Pixel-Level camera registration

State of the art

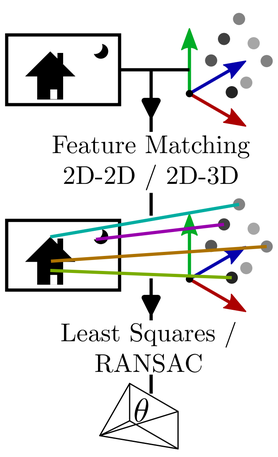


Fig. 5: Correspondence-based methods

Correspondence methods : From 2D-3D matches, one can estimate the camera model's parameters thanks to least square minimization or RANSAC. Correspondence methods usually train neural networks to embed 2D and 3D features in a common latent space and search correspondences with feature similarity. Correspondence-based methods [2, 3, 4, 5] are able to register from large viewpoint changes, but often struggle with finding pixel-level matches

Analysis-by-Synthesis methods: These approaches iteratively synthesize views of the point cloud to maximize a similarity with the real image [6, 7, 8, 9]. Analysis-by-Synthesis generally provide finer registration but optimization may fall into local minimas

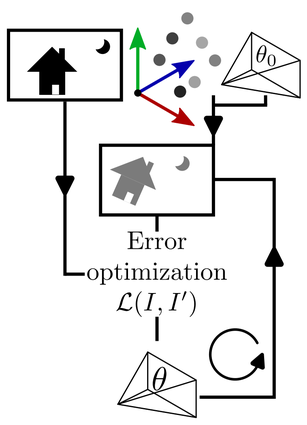


Fig. 6: Analysis-by-Synthesis methods

ReFineR

We developped ReFineR: an algorithm which uses analysis-by-synthesis for fine level registration of images and point clouds (Extrinsics only). Our method builds upon the intuition that similar objects create similar edges on an image, which can be aligned with classic optimization.

The algorithm consists of 4 main steps:

A. *Normalization*: The point cloud is converted to a normalized Gaussian Splatting model: from a first order derivation of the projection model, we compute a scaling factor which allows our method to be scale-invarient. We initialize a point splatting model by associating to each point a radius proportional to its first nearest neighbor distance.

B. *Rendering*: A differentiable view is obtained at the initial coarse camera position thanks to the Gaussian Splatting rendere, using either the point cloud normals or LiDAR reflectivity (if available).

C. *Comparison*: The Gradient Orientation Measure (GOM) is computed between the real and synthesized image as the supervision loss

D. *Optimization*: This pipeline is entirely differentiable, therefore an optimization step can be performed using gradient descent. This loops restarts from step 2 until convergence

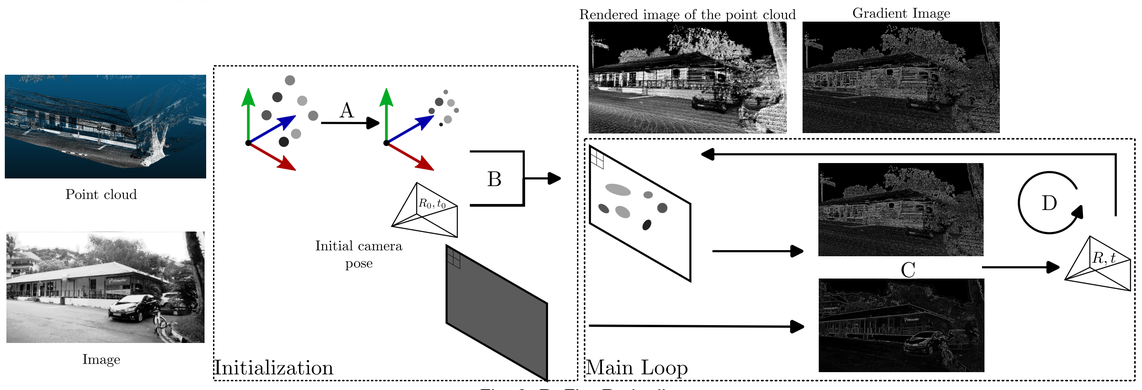


Fig. 9: ReFineR pipeline

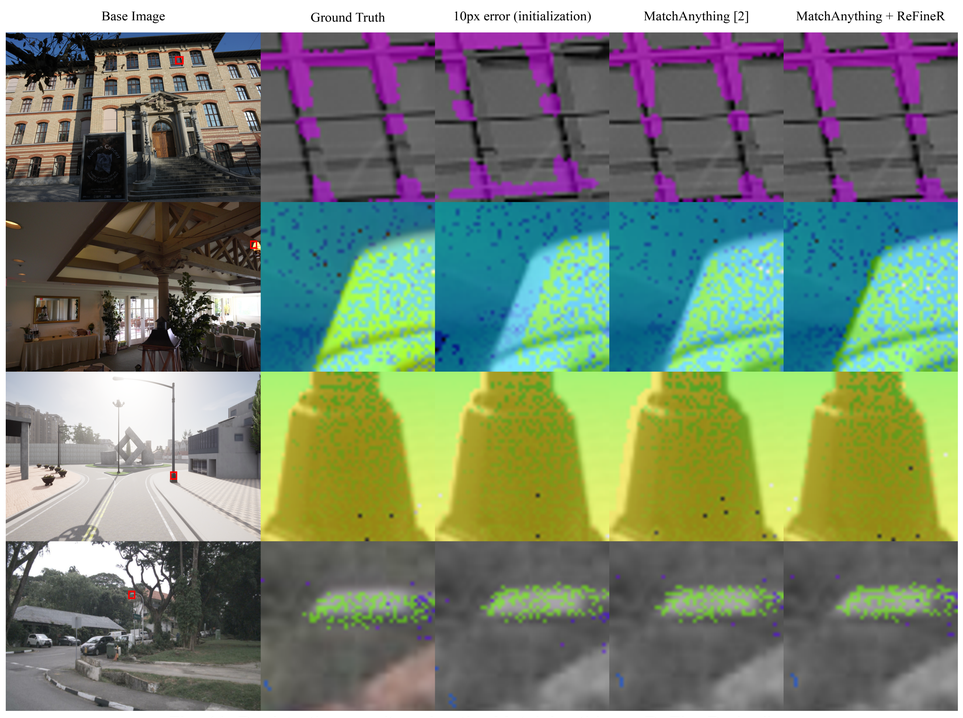


Fig. 11: Registration results using the MatchAnything + ReFineR pipeline

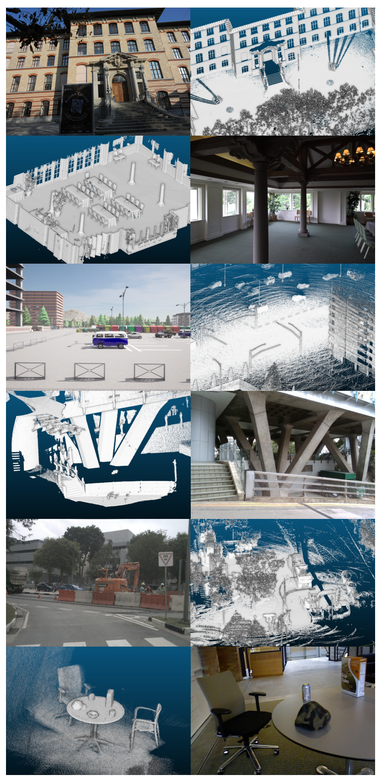


Fig. 7: Exemple images and point clouds from our evaluation dataset

Data collection

As few benchmark carry out the evaluation up to pixel level performances, we collected our own set of data from heterogeneous sources (ETH3D [10], Tanks And Temples [11], CARLA [12], Livox, nuScenes [13], RGBDScenes [14]). For Tanks And Temples and nuScenes, correspondences between 2D pixels and 3D point were manually selected to estimate our ground truth camera parameters.

Fig. 8: Point cloud registration
Left: official registration. Right: manual registration

Experiments

Setup: We randomly perturbate the ground truth camera positions by 10 pixels and observe the registration performance on our benchmark data (only rotation and translation parameters are optimized). Images are synthesized with point cloud normals or intensity. We report the results in term of percentage of images registered below a fixed pixel error threshold (pix@<threshold>).

Observations:

• Using MatchAnything as first stage and ReFineR as second, we obtained performant results accross all point cloud modality (Fig. 10 top).

• Adding LiDAR reflectivity or color intensity (when available) greatly improves the registration ability (Fig. 10 bottom)

• Even at large perturbation, MatchAnything is capable of obtaining registrations under 10px, the combination with ReFineR provides a robust pipeline (Fig. 12)

	ETH3D				Tanks and Temples				Livox				nuScenes				RGBDScenes			
	pix@ (%↑) 1	2	5	p_e(px↓)	pix@ (%↑) 2	3	5	p_e(px↓)	pix@ (%↑) 5	10	p_e(px↓)		pix@ (%↑) 3	5	10	p_e(px↓)	pix@ (%↑) 3	5	10	p_e(px↓)
One stage																				
GS-SLAM [16]	19	40	47	7.54±39.00	45	62	67	2.06±59.53	20	25	56.41±32.79		0	8	8	131.07±94.68	1	5	11	37.54±29.24
Koide *et al.* [6]	57	98	100	0.93±0.29	51	95	100	1.99±0.50	75	100	4.66±0.28		0	10	40	11.14±7.43	25	45	85	5.32±22.24
Taylor *et al.* [16]	1	1	29	6.76±5.05	3	7	23	7.25±3.48	0	75	9.09±6.69		0	0	15	22.44±46.66	0	5	55	9.07±6.65
MatchAnything [2]	35	87	100	1.12±0.61	38	87	100	2.13±0.63	75	100	4.17±0.92		4	64	100	4.72±1.19	81	100	100	1.81±0.93
ReFineR	61	70	73	0.71±3.62	71	91	98	1.93±0.94	45	95	5.17±1.68		6	35	84	6.99±2.88	12	22	81	8.07±3.00
Two stages																				
MatchAnything + GS-SLAM	32	66	73	1.30±38.55	36	74	83	2.12±38.45	25	25	65.64±30.08		0	11	17	139.64±101.83	5	5	10	35.67±31.37
MatchAnything + Koide *et al.*	57	100	100	0.95±0.25	48	96	100	2.02±0.48	50	100	4.99±0.71		0	18	68	7.99±10.92	38	60	85	3.96±27.30
MatchAnything + Taylor *et al.*	68	89	98	0.81±1.45	36	91	97	2.23±0.95	70	100	4.48±0.63		4	45	80	5.35±26.49	50	78	93	2.99±3.53
MatchAnything + MatchAnything	62	96	100	0.90±0.43	52	96	99	1.98±0.64	75	100	4.12±0.57		2	54	100	4.83±1.24	87	100	100	2.00±0.64
MatchAnything* + ReFineR	86	100	100	0.57±0.29	74	95	100	1.90±0.55	75	100	4.17±0.63		10	62	100	4.29±1.40	99	100	100	1.39±0.54

	ETH3D				Tanks and Temples				CARLA				Livox				nuScenes				RGBDScenes			
	pix@ (%↑) 1	2	5	p_e(px↓)	pix@ (%↑) 2	3	5	p_e(px↓)	pix@ (%↑) 1	2	5	p_e(px↓)	pix@ (%↑) 5	10	p_e(px↓)		pix@ (%↑) 3	5	10	p_e(px↓)	pix@ (%↑) 3	5	10	p_e(px↓)
One stage																								
GS-SLAM [15]	0	0	0	123.23±64.59	0	0	0	264.12±181.23	0	0	0	371.52±214.75	0	0	113.85±34.13		0	0	0	309.77±110.45	0	0	0	90.32±32.97
Yuan *et al.* [7]	0	2	20	10.29±9.74	1	17	52	4.39±7.71	0	2	4	15.20±41.70	25	100	5.83±1.21		0	0	5	21.10±37.53	0	0	5	26.95±26.01
Koide *et al.* [6]	0	3	25	11.10±30.86	0	0	4	29.00±67.16	2	8	28	8.82±10.73	25	75	6.61±2.60		0	0	40	12.03±12.21	0	0	5	36.19±25.14
Taylor *et al.* [16]	0	1	27	7.13±3.00	2	5	18	7.79±2.47	0	5	10	8.87±3.75	0	75	8.35±1.09		0	0	25	14.61±26.77	0	10	25	12.77±7.73
MatchAnything [2]	8	65	100	1.80±0.83	1	46	95	3.05±0.94	0	14	65	4.12±3.88	65	100	4.44±0.86		0	28	88	5.71±2.56	0	28	95	6.07±1.90
ReFineR	41	73	79	1.10±3.17	64	88	94	1.86±1.87	12	44	74	2.36±3.45	30	90	5.50±2.13		9	41	87	6.29±3.08	1	8	72	8.96±2.08
Two stages																								
MatchAnything + GS-SLAM	0	0	0	122.55±64.26	0	0	0	219.32±204.55	0	0	0	356.62±492.77	0	0	117.12±33.88		0	0	0	320.79±110.81	0	0	0	86.55±43.01
MatchAnything + Yuan *et al.*	0	1	30	8.99±48.80	0	11	55	4.56±7.63	0	2	13	12.32±17.25	50	100	4.94±1.30		0	2	14	21.78±44.35	0	0	5	32.08±20.41
MatchAnything + Koide *et al.*	1	14	45	5.90±17.58	0	5	7	22.25±69.57	0	5	29	7.70±22.69	0	85	7.16±1.92		0	13	54	8.38±46.37	0	12	37	37.68±26.45
MatchAnything + Taylor *et al.*	21	74	96	1.48±1.85	12	65	98	2.72±0.81	0	9	51	4.97±3.77	25	100	5.69±1.17		0	31	80	6.62±2.65	1	27	63	7.31±6.01
MatchAnything + MatchAnything	14	70	99	1.58±0.85	8	71	99	2.63±0.77	3	24	66	3.38±4.31	50	100	5.13±0.76		0	32	98	5.70±1.93	9	37	91	5.58±2.51
MatchAnything* + ReFineR	58	96	100	0.90±0.40	70	95	100	1.76±0.51	16	68	91	1.54±3.04	55	100	4.94±0.94		6	54	97	4.90±2.75	13	43	92	5.36±4.25

Fig. 10: Registration experiments at 10px using LiDAR intensity (top) and point cloud normals (bottom)

	CARLA										
	pix@ (%↑) 1	2	3	5	10	20	50	100	p_e(px↓)	rot (↓)	transl (cm↓)
0px initialization											
MatchAnything [2]	3	26	44	85	98	100	100	100	3.20±2.60	0.14±0.08	33.94±27.77
ReFineR (Ours)	20	76	92	100	100	100	100	100	1.47±0.79	0.06±0.04	2.99±1.74
MatchAnything* + ReFineR	16	71	83	95	96	99	100	100	1.49±2.72	0.05±0.09	3.03±6.32
5px initialization											
MatchAnything [2]	2	17	39	78	95	99	100	100	3.35±3.20	0.13±0.12	32.50±32.13
ReFineR (Ours)	11	63	82	95	99	100	100	100	1.68±1.72	0.06±0.07	3.29±3.82
MatchAnything* + ReFineR	16	66	84	96	99	100	100	100	1.49±1.59	0.06±0.06	3.05±3.72
10px initialization											
MatchAnything [2]	0	14	30	65	90	100	100	100	4.12±3.88	0.15±0.17	37.95±53.56
ReFineR (Ours)	12	44	57	74	92	100	100	100	2.36±3.45	0.08±0.25	4.44±7.52
MatchAnything* + ReFineR	16	68	81	91	97	99	100	100	1.54±3.04	0.06±0.09	3.05±6.39
20px initialization											
MatchAnything [2]	0	11	19	49	85	96	100	100	5.13±4.94	0.19±0.24	43.47±66.34
ReFineR (Ours)	2	14	18	21	26	82	100	100	16.67±7.70	0.94±0.64	26.50±16.88
MatchAnything* + ReFineR	18	66	82	90	94	98	100	100	1.54±4.19	0.06±0.22	3.06±8.75
50px initialization											
MatchAnything [2]	0	8	18	34	65	93	98	99	6.79±14.99	0.23±0.32	50.53±79.29
ReFineR (Ours)	0	0	0	0	0	0	56	100	49.60±6.50	3.63±1.19	82.69±29.06
MatchAnything* + ReFineR	17	54	73	80	92	96	99	100	1.64±9.36	0.06±0.29	3.35±20.37
100px initialization											
MatchAnything [2]	1	3	9	31	55	77	95	100	8.32±15.98	0.28±0.87	81.66±202.21
ReFineR (Ours)	0	0	0	0	0	0	0	52	99.66±13.05	7.43±2.28	170.25±56.97
MatchAnything* + ReFineR	12	52	60	69	80	89	95	99	1.98±20.24	0.07±0.65	4.29±41.36

Fig. 12: Registration at different perturbation levels

Perspectives

Following this work, several aspects will be studied more in details

• Adapting ReFineR for camera intrinsics and distortion. For now, only gradients with respect to the camera position were implemented in the custom Gaussian Splatting rendered by Matsuki et al. [15]. We will propose to add gradients w.r.t. other camera parameters.

• Registration at larger scale and without pose prior. In regard of the performaces of MatchAnything for large perturbation, a comprehensive registration pipeline could be implemented by matching an image with known rendered views of the scene.

• New data. Having seen that most datasets lack pixel level registration. We will propose our own dataset of pixel-calibrated and registered images with diverse cameras and LiDAR sensors.

[1] 3D Gaussian Splatting for Real-Time Radiance Field Rendering, Kerbl *et al.* 2023
[2] MatchAnything: Universal cross-modality image matching with large-scale pre-training, He *et al* 2025
[3] UniCorrn: Unified Correspondence Transformer Across 2D and 3D, Goswami *et al.* 2026
[4] CoFiI2P: Coarse-to-Fine Correspondences for Image-to-Point Cloud Registration, Kang *et al.* 2023
[5] Corri2p: Deep image-to-point cloud registration via dense correspondence, Ren et al. 2022
[6] General, Single-shot, Target-less, and Automatic LiDAR-Camera Extrinsic Calibration Toolbox, Koide et al. 2023
[7] Pixel-Level Extrinsic Self Calibration of High Resolution LiDAR and Camera in Targetless Environments, Yuan et al. 2021
[8] LCCNet: LiDAR and Camera Self-Calibration using Cost Volume Network, Lv et al. 2021
[9] CalibNet: Self-Supervised Extrinsic Calibration using 3D Spatial Transformer Networks, Iyer et al. 2018
[10] A multi-view stereo benchmark with high-resolution images and multi-camera videos, Schöps et al. 2017
[11] Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction, Knapitsch et al. 2017
[12] CARLA: An open urban driving simulator, Dosovitskiy et al. 2017
[13] nuScenes: a multimodal dataset for autonomous driving, Caesar et al. 2020
[14] Unsupervised feature learning for 3d scene labeling, Lai et al. 2014
[15] Gaussian splatting SLAM, Matsuki et al. 2023
[16] Multi-Modal Sensor Calibration Using a Gradient Orientation Measure, Taylor et al. 2014