

RGB-based Simultaneous Localization and Mapping (SLAM): State of the Art - Abschlusspräsentation

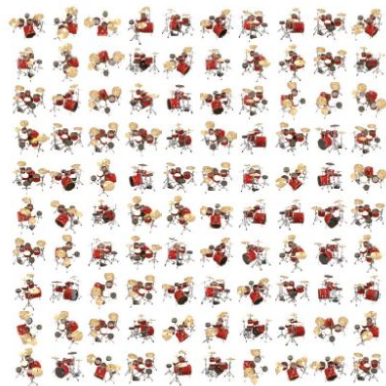
Johannes Decker
01.07.2025

Neural Radiance Fields (NeRFs)

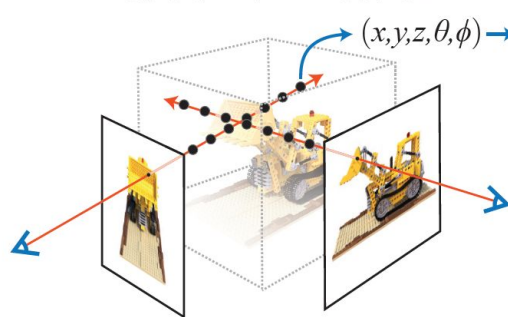
Input Images

Optimize NeRF

Render new views

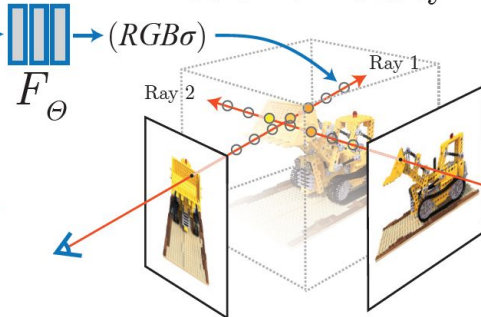


5D Input
Position + Direction



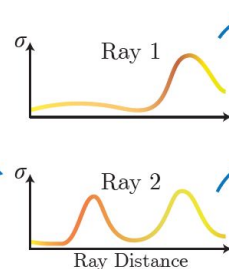
(a)

Output
Color + Density



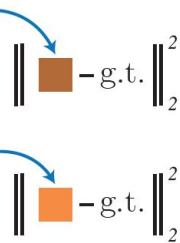
(b)

Volume
Rendering



(c)

Rendering
Loss

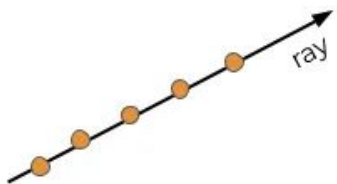


(d)

NeRFs vs 3DGS

NeRF

Gaussian Splatting



REFINED

POINT
CLOUD

GAUSSIANS

Performance Vergleich - Metriken für die Bildqualität

- **SSIM (Structural Similarity Index \uparrow)** bewertet die strukturelle Ähnlichkeit zweier Bilder, und berücksichtigt Helligkeit Kontrast – ähnlich wie das menschliche Sehen.
- **PSNR (Peak Signal-to-Noise Ratio \uparrow)** misst die Differenz (Fehler) zwischen den Pixelwerten zweier Bilder, und stellt eine technische, nicht wahrnehmungsbasierte Metrik dar.
- **LPIPS (Learned Perceptual Image Patch Similarity \downarrow)** verwendet ein neuronales Netz (z. B. VGG oder AlexNet), um zu bewerten, wie unterschiedlich zwei Bilder wahrgenommen werden.

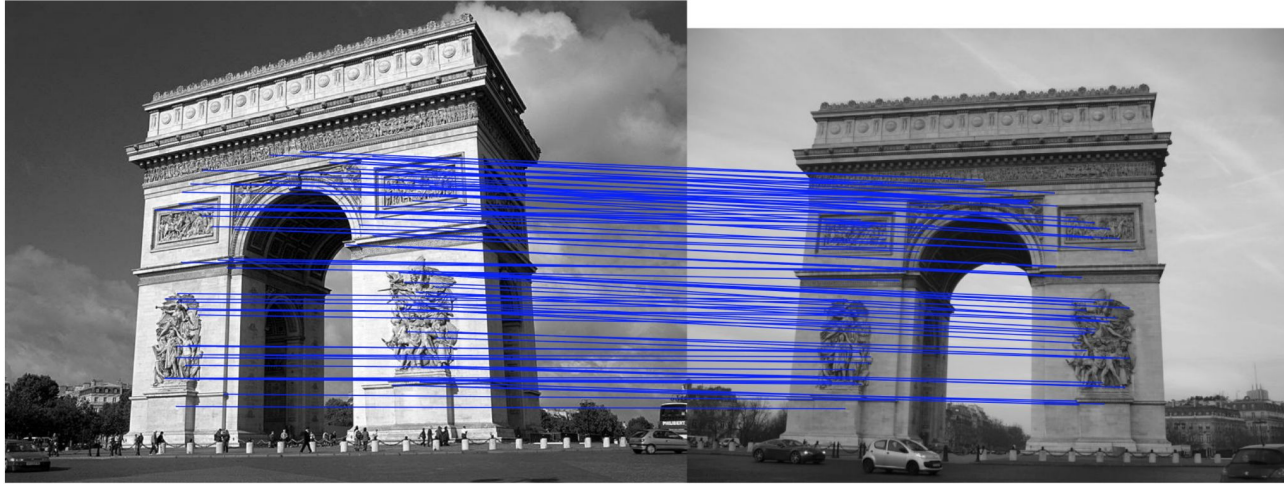
Ergebnisse GIORIE-SLAM

Dataset	Subset	PSNR \uparrow	SSI \uparrow	LPIPS \downarrow	Computation time [h:min:sec] \downarrow	Nr. frames
Replica	office0	33.939	0.972	0.113	3:51:17	2000
	office1	36.084	0.986	0.089	3:13:38	2000
	office2	26.446	0.957	0.193	4:24:32	2000
	office3	26.402	0.951	0.170	3:47:48	2000
	office4	28.751	0.955	0.214	4:07:54	2000
	room0	27.015	0.945	0.174	3:36:58	2000
	room1	28.201	0.944	0.190	4:31:33	2000
	room2	28.484	0.945	0.192	4:08:31	2000
	Avg.	29.415	0.957	0.167	03:57:46	—
TUM RGB-D	desk	16.845	0.715	0.652	1:12:27	595
	desk2	15.578	0.614	0.764	3:33:32	639
	room	15.798	0.607	0.746	8:28:37	1360
	xyz	21.014	0.855	0.357	1:20:02	3666
	household	16.923	0.622	0.745	2:39:32	2509
	Avg.	17.232	0.683	0.653	03:26:50	—

Performance Vergleich - Metriken für die Geometrie

- Accuracy (Acc. [cm] ↓): Wie nahe liegen die rekonstruierten Punkte an den Punkten der Ground-Truth-Oberfläche?
- Completeness (Comp. [cm] ↓): Wie nahe liegen die Ground-Truth-Punkte an irgendeinem rekonstruierten Punkt? Wurden alle Teile des Objekts erfasst?
- Completeness Ratio (Comp. Rat [%] ↑): Wie viel Prozent der Ground-Truth-Punkte liegen innerhalb eines kleinen Abstands (GLORIE-SLAM: 2cm; HI-SLAM2: 5cm) zu einem rekonstruierten Punkt?

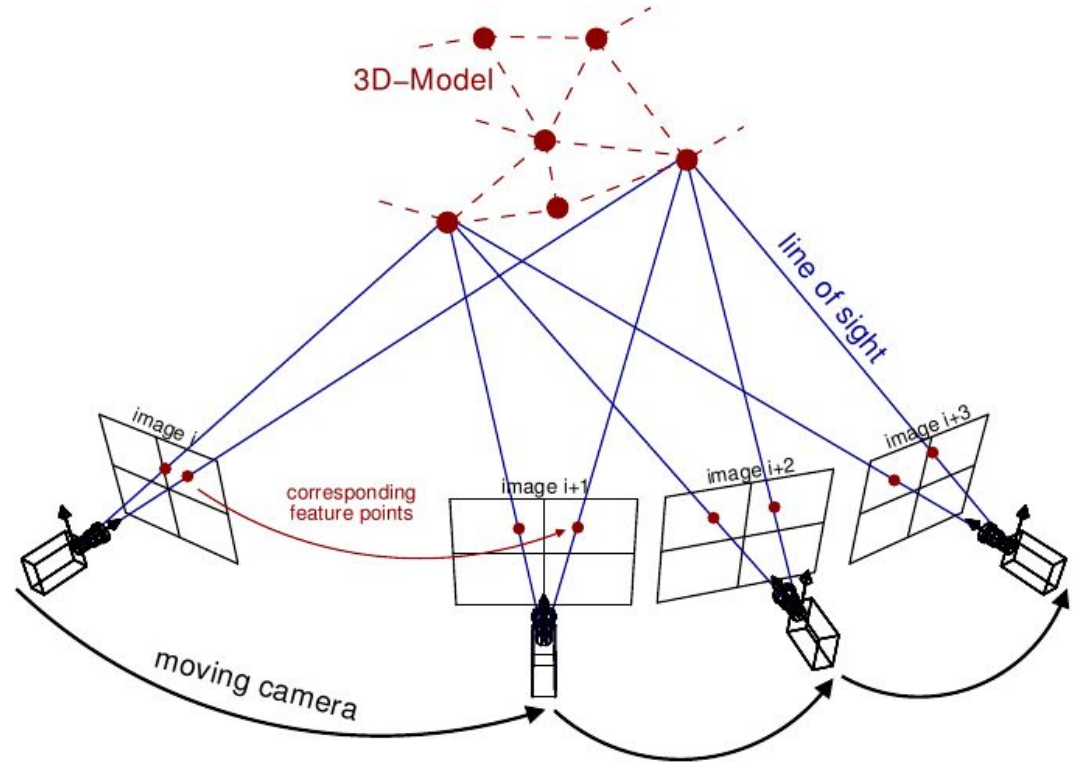
Datenaufbereitung - Feature Detection and Matching



Erkennen und Zuordnen von Schlüssel-Punkten über mehrere Bilder hinweg

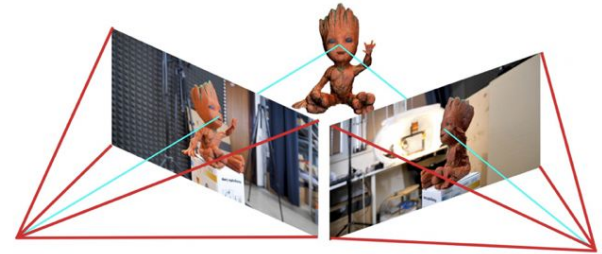
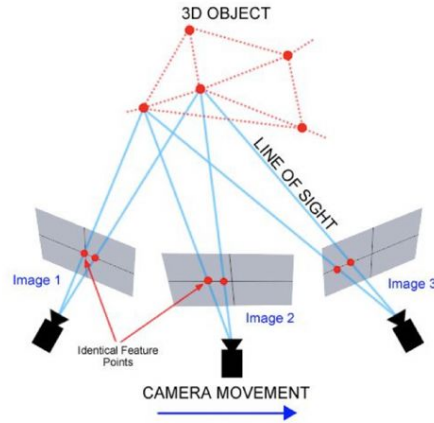
Datenaufbereitung - Kamera-Posen-Schätzung

- Schätzen relativer Kamerapositionen (Rotation + Translation) mithilfe der Epipolargeometrie
- Schrittweises hinzunehmen weitere Bilder, um eine Kamera-Trajektorie zu erstellen (Verbesserung der Posen bzw. Trajektorie durch Bundle Adjustment)
- Lückenhafte Punktwolke entsteht



Datenaufbereitung - Erzeugen einer dichten Punktwolke

- Berechnung einer dichten Tiefenkarte für jedes Bild unter Verwendung konsistenter, nahegelegener Ansichten
- Zusammenführen der Tiefenkarten aus verschiedenen Ansichten zu einer einzigen dichten Punktwolke

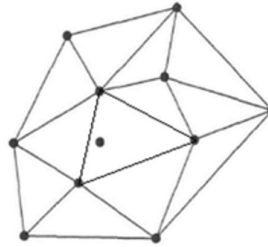


Datenaufbereitung - Generierung eines Netzes

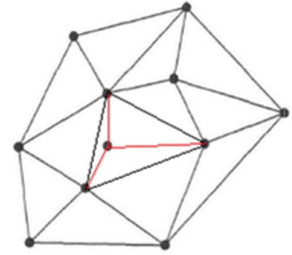
Konvertieren der dichten Punktwolke in ein Netz aus Dreiecken, z.B. durch **Delaunay-Triangulation**:

=> Der Umkreis eines Dreiecks des Netzes darf keine weiteren Punkte enthalten

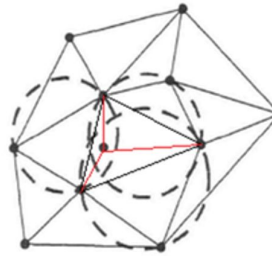
=> Dadurch wird der kleinste Innenwinkel über alle Dreiecke maximiert



(a) Insert a new point in Delaunay triangle



(b) Connect the new point to the three surrounding vertices



(c) Use Local Optimization Procedure to optimize the triangular network



(d) Generate the new Delaunay triangle

Ergebnisse HI-SLAM2

Dataset	Subset	PSNR \uparrow	SSI \uparrow	LPIPS \downarrow	Acc.[cm] \downarrow	Comp.[cm] \downarrow	Comp. Rat[%] \uparrow	Computation time [min:sec] \downarrow	Nr. of frames
Replica	office0	42.641	0.984	0.021	0.014	0.021	0.894	2:42	2000
	office1	43.621	0.984	0.021	0.011	0.032	0.856	2:27	2000
	office2	36.812	0.969	0.042	0.020	0.044	0.864	2:52	2000
	office3	36.867	0.969	0.036	0.018	0.040	0.878	2:32	2000
	office4	38.524	0.972	0.043	0.019	0.044	0.853	2:55	2000
	room0	35.231	0.959	0.041	0.015	0.034	0.902	2:29	2000
	room1	36.655	0.965	0.047	0.013	0.031	0.888	2:56	2000
	room2	38.204	0.973	0.037	0.018	0.038	0.886	2:50	2000
	Avg.	38.569	0.972	0.036	0.016	0.035	0.878	2:43	-
TUM RGB-D	desk	21.210	0.797	0.197	1.566	10.363	0.002	20:36	595
	desk2	21.132	0.794	0.237	1.214	13.210	0.019	21:10	639
	room	20.923	0.773	0.266	1.248	16.867	0.014	26:47	1360
	xyz	24.207	0.819	0.133	6.559	171.178	-	18:10	3666
	household	21.921	0.824	0.186	1.248	13.268	0.046	11:56	2509
	Avg.	21.878	0.801	0.204	2.367	44.977	0.017	19:44	-
HM	first	15.181	0.640	0.610	2.393	788.925	0.015	10:44	1857
	second	15.846	0.667	0.559	7.010	1647.933	0.016	10:53	2616
	Avg.	15.514	0.654	0.584	4.702	1218.429	0.016	10:49	-

HI-SLAM2 basierte Rendering Videos von autonomen Modellauto der Hochschule München

Quellen

<https://www.researchgate.net/profile/Sjoerd-Van-Riel-2/publication/303824023/figure/fig3/AS:370256326479881@1465287395641/Structure-from-Motion-SfM-photogrammetric-principle-Source-Theia-sfmorg-2016.png>

<https://www.cs.toronto.edu/~fidler/slides/2021Winter/CSC420/lecture7.pdf>

<https://www.computer.org/csdl/journal/tp/2021/10/09076298/1jeCY6FnOcU>

<https://pyimagesearch.com/2024/12/09/3d-gaussian-splatting-vs-nerf-the-end-game-of-3d-reconstruction/>

<https://medium.com/data-science/a-comprehensive-overview-of-gaussian-splatting-e7d570081362>

[NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis](#)

<https://www.ailabs.tw/wp-content/uploads/2019/06/image15.png>

<https://levelup.gitconnected.com/facemask-a-real-time-face-morphing-tool-5b343591a237>

<https://doi.org/10.1007/s00366-020-01145-1>

Fragen ?