Improving RGB-D Point Cloud Registration by Learning Multi-scale Local Linear Transformation

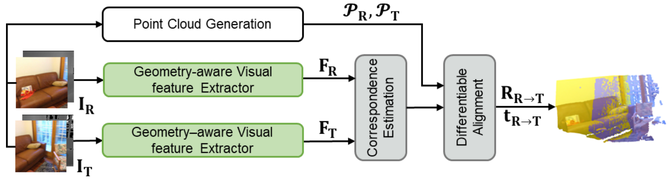
Ziming Wang[1 *], Xiaoliang Huo[1 *], Zhenghao Chen[2], Jing Zhang[1], Lu Sheng[1], and Dong Xu[3]

[1] School of Software, Beihang University
[2] School of Electrical and Information Engineering, The University of Sydney
[3] Department of Computer Science, The University of Hong Kong

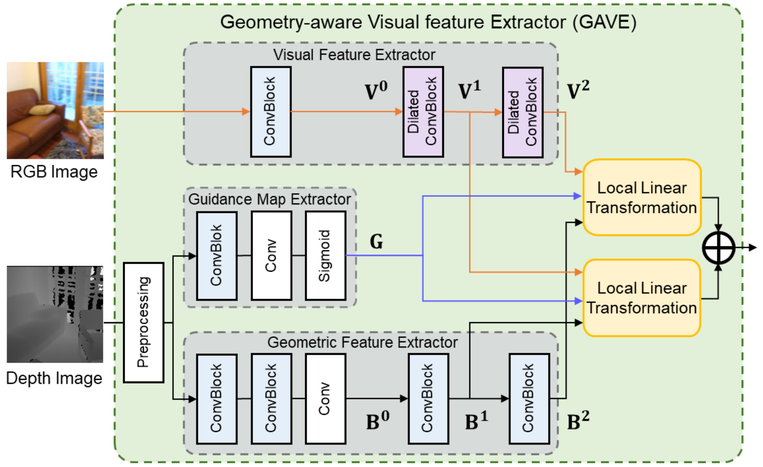
ECCV TEL AVIV 2022

Introduction

- We introduce a **Geometry-Aware Visual Feature Extractor (GAVE)** to generate distinctive but comprehensive geometric-visual features for point cloud registration.
- We propose a **Local Linear Transformation (LLT)** module , which employs the Bilateral Grid and an edge-aware guidance map to generate our content-aware linear coefficient.
- By collaborate multi-scale LLT modules in GAVE, we observe a significant performance enhancement.

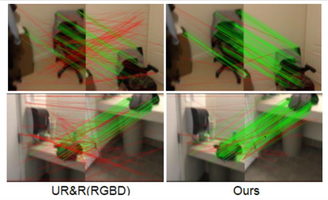


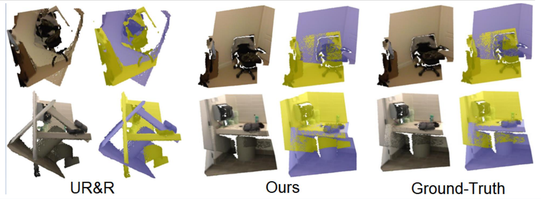
Network Architecture



Experimental Results

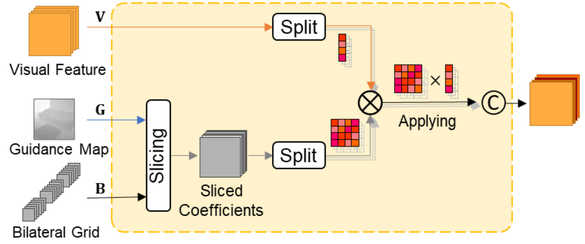
Methods	Train Set	Sup	Features Visual	Features 3D	Rotation 5	Rotation 10	Rotation 45	Translation 5	Translation 10	Translation 25	Chamfer 1	Chamfer 5	Chamfer 10
SIFT [27]	N/A		✓		55.2	75.7	89.2	17.7	44.5	79.8	38.1	70.6	78.3
SuperPoint [14]	N/A		✓		65.5	86.9	96.6	21.2	51.7	88.0	45.7	81.1	88.2
FCGF [10]	N/A			✓	70.2	87.7	96.2	27.5	58.3	82.9	52.0	78.0	83.7
BYOC [16]	3D Match		✓	✓	66.5	85.2	97.8	30.7	57.6	88.9	54.1	82.8	89.5
DGR [8]	3D Match	✓		✓	81.1	89.3	94.8	54.5	76.2	88.7	70.5	85.5	89.0
3D MV Reg [19]	3D Match	✓		✓	87.7	93.2	97.0	69.0	83.1	91.8	78.9	89.2	91.8
UR&R [15]	3D Match		✓		87.6	93.1	98.3	69.2	84.0	93.8	79.7	91.3	94.0
UR&R (RGB-D) [15]	3D Match		✓	✓	87.6	93.7	98.8	67.5	83.8	94.6	78.6	91.7	94.6
Ours	3D Match		✓	✓	**93.4**	**96.5**	**98.8**	**76.9**	**90.2**	**96.7**	**86.4**	**95.1**	**96.8**

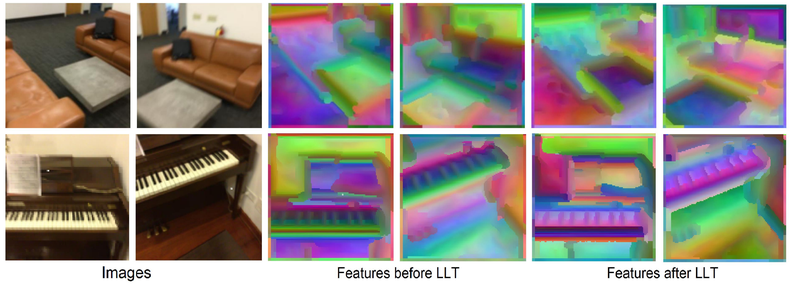




Local Linear Transformation (LLT)

- We progressively apply the local linear transformation to learn the visual-geometric features in a multi-scale manner.
- The slicing operation is performed between a **Bilateral Grid** and the **guidance map**.
- The final local linear transformation can be obtained by first using a point-wise transformation and then using a channel-wise concatenation





Ablation Study

- **Component effectiveness**
  - MS: multi-scale strategy
  - DC: dilated convolutions in the visual feature extractor

MS	DC	LLT	Rotation Accuracy 5	Rotation Accuracy 10	Rotation Accuracy 45	Rotation Error Mean	Rotation Error Med.	Translation Accuracy 5	Translation Accuracy 10	Translation Accuracy 25	Translation Error Mean	Translation Error Med.	Chamfer Accuracy 1	Chamfer Accuracy 5	Chamfer Accuracy 10	Chamfer Error Mean	Chamfer Error Med.
			88.4	94.2	98.6	3.8	1.1	67.3	83.8	94.5	8.5	3.0	78.9	91.7	94.6	6.5	0.2
✓			88.5	94.4	98.6	3.8	1.1	68.1	84.5	94.8	8.3	3.0	79.5	92.1	94.9	6.3	0.2
✓	✓		90.4	95.0	98.6	3.6	1.0	70.8	86.5	95.3	8.1	2.8	81.8	93.1	95.4	6.2	0.2
✓	✓	✓	**93.4**	**96.5**	**98.8**	**3.0**	**0.9**	**76.9**	**90.2**	**96.7**	**6.4**	**2.4**	**86.4**	**95.1**	**96.8**	**5.3**	**0.1**

- **Optimized in supervised manner**

Methods	Rotation Accuracy 5	Rotation Accuracy 10	Rotation Accuracy 45	Rotation Error Mean	Rotation Error Med.	Translation Accuracy 5	Translation Accuracy 10	Translation Accuracy 25	Translation Error Mean	Translation Error Med.	Chamfer Accuracy 1	Chamfer Accuracy 5	Chamfer Accuracy 10	Chamfer Error Mean	Chamfer Error Med.
UR&R (RGB-D)	92.3	95.3	98.2	3.8	0.8	77.6	89.4	95.5	7.8	2.3	86.1	94.0	95.6	6.7	0.1
Ours	**96.5**	**97.8**	**98.8**	**2.7**	**0.8**	**83.8**	**93.8**	**97.6**	**5.8**	**2.0**	**91.2**	**96.7**	**97.6**	**4.8**	**0.1**