Accurate Registration of Cross-Modality Geometry via Consistent Clustering

Mingyang Zhao[1]

Joint work with

Xiaoshui Huang[2], Jingen Jiang[3], Luntian Mou[4], Lei Ma[5], and Dong-Ming Yan[6]

中国科学院香港创新研究院人工智能与机器人创新中心
Centre for Artificial Intelligence and Robotics
Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences

上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

SHANDONG UNIVERSITY 1901

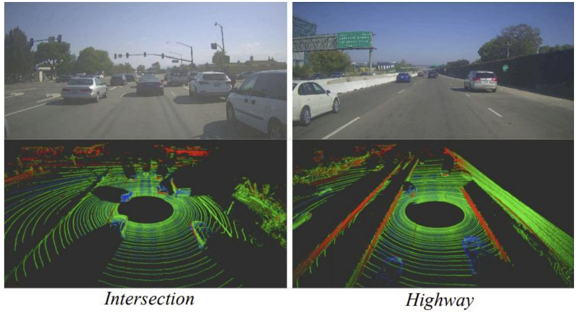
BEIJING UNIVERSITY OF TECHNOLOGY 北京工业大学 1960

PEKING UNIVERSITY 1898

CASIA
中国科学院自动化研究所
INSTITUTE OF AUTOMATION
CHINESE ACADEMY OF SCIENCES

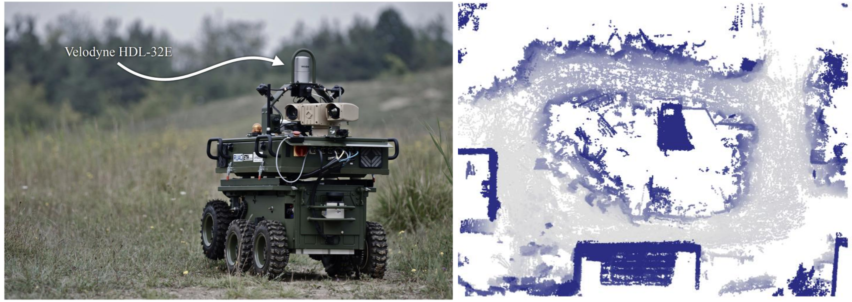
APPLICATIONS - SHAPE REGISTRATION

autonomous driving



[Lu et al. 2019]

robotics



[Pomerleau et al. 2015]

dynamic reconstruction

[Yao et al. 2024]

shape completion

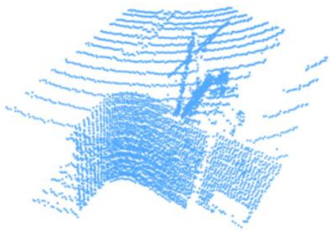
[Halimi et al. 2020]

MOTIVATION - CROSS-MODALITY REGISTRATION

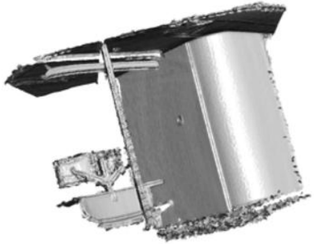
1. Cross-modality data having different representations, e.g., densities, textures, ...

2. Develop a holistic framework to deal with cross-modality registration

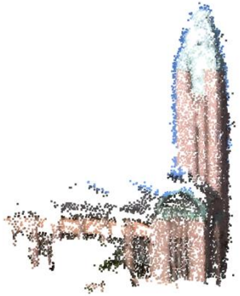
3. Consider three common modalities including LiDAR, Kinect, and MVS



LiDAR



Kinect

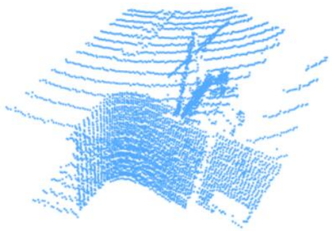


MVS

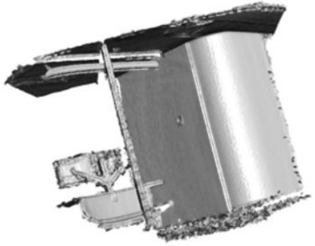
MOTIVATION - CROSS-MODALITY REGISTRATION

Observations: different representations BUT **similar geometric structures**

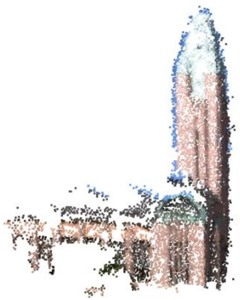
How to define geometric structures?



LiDAR



Kinect



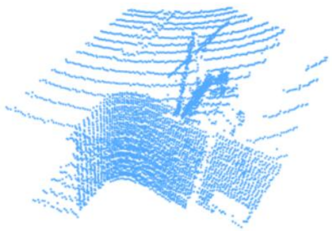
MVS

MOTIVATION - CROSS-MODALITY REGISTRATION

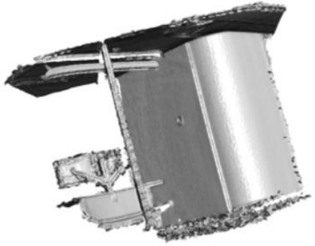
Observations: different representations BUT **similar geometric structures**

How to define geometric structures?

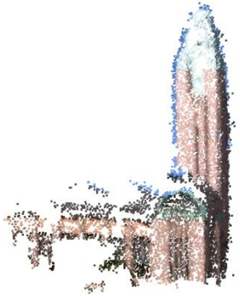
Clustering



LiDAR

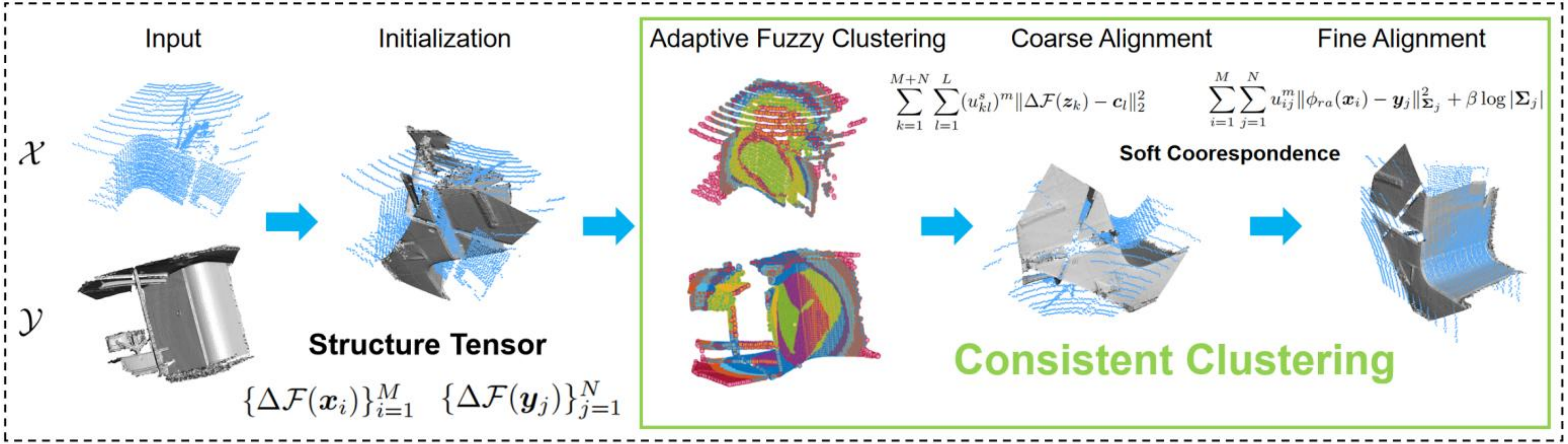


Kinect



MVS

METHOD - UNSUPERVISED CLUSTERING



Formulate cross-modality registration as **an unsupervised clustering problem**

METHOD - UNSUPERVISED CLUSTERING

Fuzzy clustering to discover similar structures

$$\boxed{\min_{\mathbf{U}_{\mathbf{s}},\mathcal{C}} J(\mathbf{U}_{\mathbf{s}},\mathcal{C}) = \sum_{k=1}^{M+N}\sum_{l=1}^{L}(u_{kl}^{s})^{m}\|\Delta\mathcal{F}(\boldsymbol{z}_k)-\boldsymbol{c}_l\|_2^2}$$

Clustering centroids

$$\boldsymbol{c}_l = \frac{\sum_{k=1}^{M+N} u_{kl}^{m}\Delta\mathcal{F}(\boldsymbol{z}_k)}{\sum_{k=1}^{M+N} u_{kl}^{m}}, \quad u_{kl}^{s} = \frac{\left(\frac{1}{\|\Delta\mathcal{F}(\boldsymbol{z}_k)-\boldsymbol{c}_l\|_2^2}\right)^{\frac{1}{m-1}}}{\sum_{l'=1}^{L}\left(\frac{1}{\|\Delta\mathcal{F}(\boldsymbol{z}_k)-\boldsymbol{c}_{l'}\|_2^2}\right)^{\frac{1}{m-1}}}$$

Fuzzy degree member

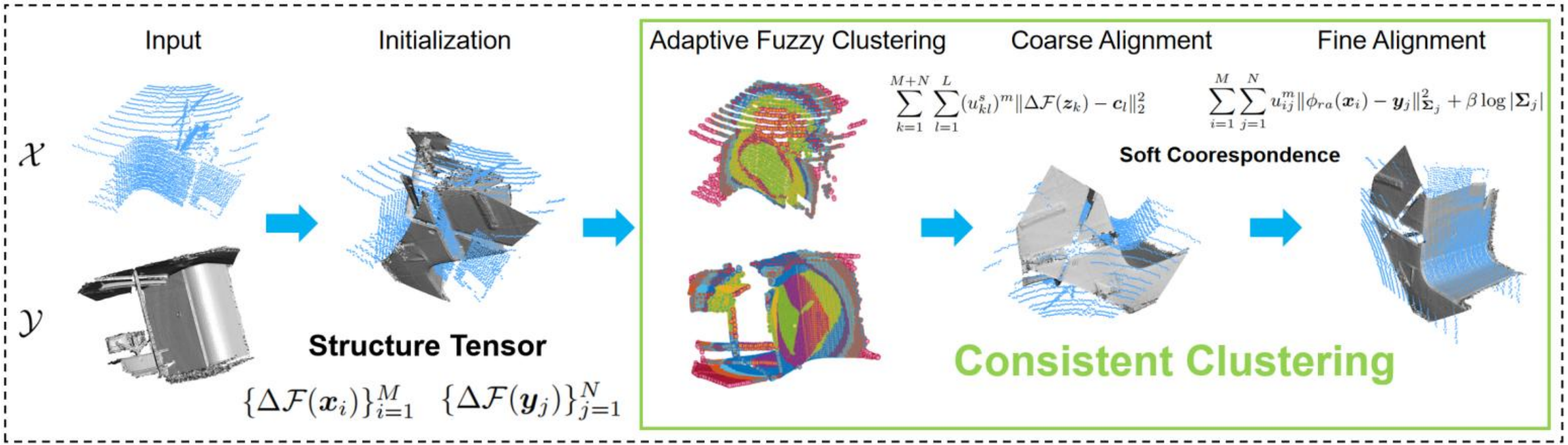
Moment difference

$$\Delta\mathcal{F}(\boldsymbol{x}) = (\Delta\mathcal{F}_1(\boldsymbol{x}), \Delta\mathcal{F}_2(\boldsymbol{x}), \cdots, \Delta\mathcal{F}_{n-1}(\boldsymbol{x})) \in \mathbb{R}^{n-1}$$

Multi-scale moment

$$\mathcal{F}_{r_i}(\boldsymbol{x}) = (I_{r_i1}, I_{r_i2}, I_{r_i3}) \quad I_1 = \frac{\lambda_3}{\lambda_1}, \quad I_2 = \sqrt{\lambda_1\cdot\lambda_2\cdot\lambda_3}, \quad I_3 = -\sum_{i=1}^{3}\lambda_i \ln \lambda_i$$

METHOD - COARSE REGISTRATION



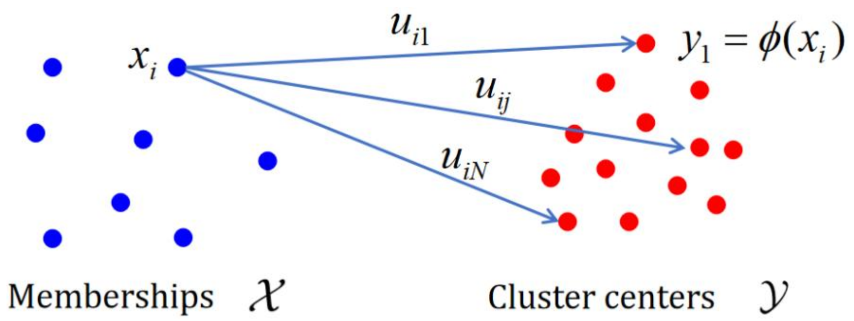
Coarse registration

$$\boxed{\phi_{ca} = \min_{\{\phi_l\}_{l=1}^{L}} \|\phi_l(\mathcal{X}) - \mathcal{Y}\|_2,\ s.t.\ \phi_l = \min_{\phi} \|\phi(\tilde{\mathcal{X}}) - \tilde{\mathcal{Y}}\|_2}$$

Fuzzy clustering to discover similar structures ⇨

$$\tilde{\mathcal{X}} = \{x_1, x_2, x_3\} \subset \mathcal{X},\ \tilde{\mathcal{Y}} = \{y_1, y_2, y_3\} \subset \mathcal{Y}$$

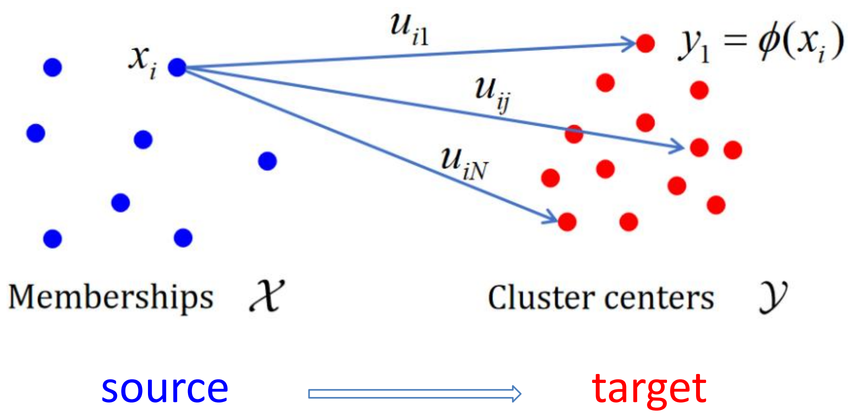
METHOD - FINE REGISTRATION



source $\Rightarrow$ target

Formulate fine point-wise alignment still as an unsupervised clustering process

METHOD - FINE REGISTRATION



$$\min_{\phi_{ra},\sigma^2} J(\phi_{ra},\sigma^2) = \sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij}^{m} \|\phi_{ra}(\boldsymbol{x}_i) - \boldsymbol{y}_j\|^2_{\boldsymbol{\Sigma}_j} + \beta \log|\boldsymbol{\Sigma}_j|$$

METHOD - FINE REGISTRATION

Objective function

$$\boxed{\min_{\phi_{ra},\sigma^2} J(\phi_{ra},\sigma^2) = \sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij}^{m} \|\phi_{ra}(\boldsymbol{x}_i) - \boldsymbol{y}_j\|_{\boldsymbol{\Sigma}_j}^2 + \beta \log|\boldsymbol{\Sigma}_j|}$$

$$\phi_{ra}(\boldsymbol{x}_i) = \mathbf{R}\boldsymbol{x}_i + \mathbf{t}$$

$$\|\phi_{ra}(\boldsymbol{x}_i) - \boldsymbol{y}_j\|_{\boldsymbol{\Sigma}_j}^2 = (\phi_{ra}(\boldsymbol{x}_i) - \boldsymbol{y}_j)^T \boldsymbol{\Sigma}_j^{-1} (\phi_{ra}(\boldsymbol{x}_i) - \boldsymbol{y}_j)$$

Fuzzy degree member

$$u_{ij}^{d} = \frac{\left(\frac{1}{\|\phi_{ra}(\boldsymbol{x}_i) - \boldsymbol{y}_j\|_{\boldsymbol{\Sigma}_j}^2}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{N}\left(\frac{1}{\|\phi_{ra}(\boldsymbol{x}_i) - \boldsymbol{y}_k\|_{\boldsymbol{\Sigma}_k}^2}\right)^{\frac{1}{m-1}}} = \frac{1}{\sum_{k=1}^{N}\left(\frac{\|\phi_{ra}(\boldsymbol{x}_i) - \boldsymbol{y}_j\|_{\boldsymbol{\Sigma}_j}}{\|\phi_{ra}(\boldsymbol{x}_i) - \boldsymbol{y}_k\|_{\boldsymbol{\Sigma}_k}}\right)^{\frac{2}{m-1}}}$$

METHOD - FINE REGISTRATION

Objective function

$$\boxed{\min_{\phi_{ra},\sigma^2} J(\phi_{ra},\sigma^2) = \sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij}^{m} \|\phi_{ra}(\boldsymbol{x}_i) - \boldsymbol{y}_j\|_{\boldsymbol{\Sigma}_j}^2 + \beta \log |\boldsymbol{\Sigma}_j|}$$

Theorem 2. *The upper bound of the fuzzy registration residual (Eq. 4) is minimized by the following equations:*

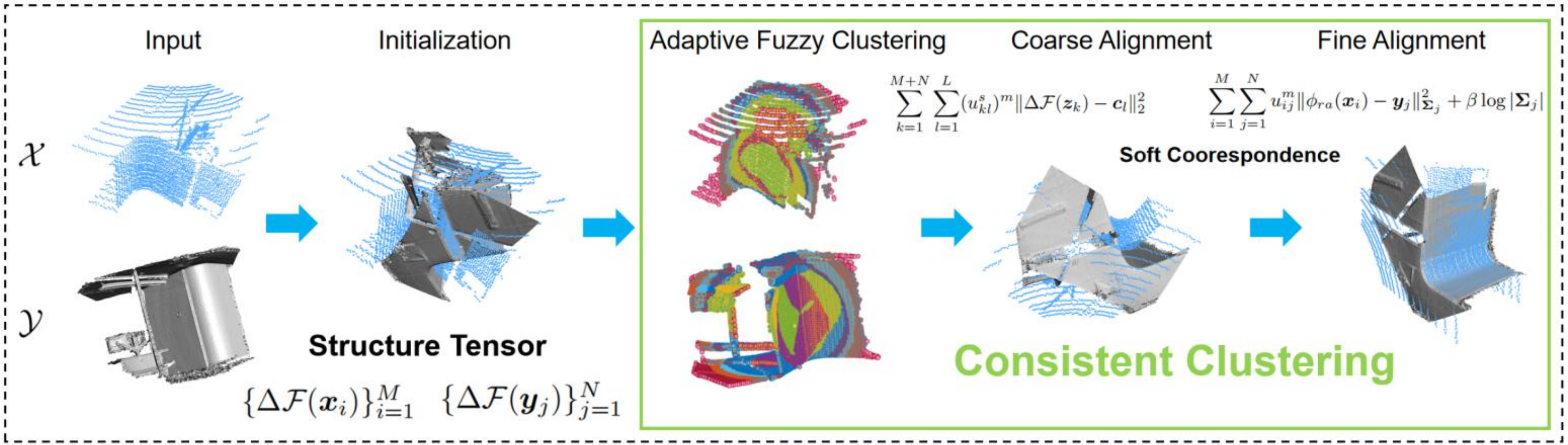
$$\hat{\mathbf{R}} = \boldsymbol{\Phi}\mathbf{d}(1,\cdots,1,|\boldsymbol{\Phi}\boldsymbol{\Psi}^T|)\boldsymbol{\Psi}^T, \quad \hat{\mathbf{t}} = \bar{\boldsymbol{y}} - \hat{\mathbf{R}}\bar{\boldsymbol{x}},$$

$$\hat{\sigma}^2 = \frac{1}{3S}\mathrm{tr}(\bar{\mathbf{X}}\mathbf{d}(\mathbf{U}\mathbf{1})\bar{\mathbf{X}}^{\mathbf{T}} - \mathbf{2}\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Psi}\hat{\mathbf{R}}^{\mathbf{T}} + \bar{\mathbf{Y}}\mathbf{d}(\mathbf{U}^{\mathbf{T}}\mathbf{1})\bar{\mathbf{Y}}^{\mathbf{T}}),$$

where $\mathbf{X} = [\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m] \in \mathbb{R}^{3\times M}$, $\mathbf{Y} = [\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n] \in \mathbb{R}^{3\times N}$, $S = \sum_{i,j=1}^{M,N} u_{ij}$, $\bar{\mathbf{x}} = \frac{1}{S}\mathbf{X}\mathbf{U}\mathbf{1}$, $\bar{\boldsymbol{y}} = \frac{1}{S}\mathbf{Y}\mathbf{U}^{\mathbf{T}}\mathbf{1}$, $\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Psi}^{\mathbf{T}} = svd((\mathbf{Y}-\bar{\boldsymbol{y}}\mathbf{1})\mathbf{U}^{\mathbf{T}}(\mathbf{X}-\bar{\boldsymbol{x}}\mathbf{1})^{\mathbf{T}}) = svd(\bar{\mathbf{Y}}\mathbf{U}^{\mathbf{T}}\bar{\mathbf{X}}^{\mathbf{T}})$, $\mathbf{1}$ *is the corresponding full one column vector.*

Analytical closed-form solution (all unknowns) ⇒ Fast alternative optimization (coordinate descent)

METHOD – CROSS-MODALITY REGISTRATION



discover structure similarity | coarse | fine

Formulate cross-modality registration as **an unsupervised clustering problem**

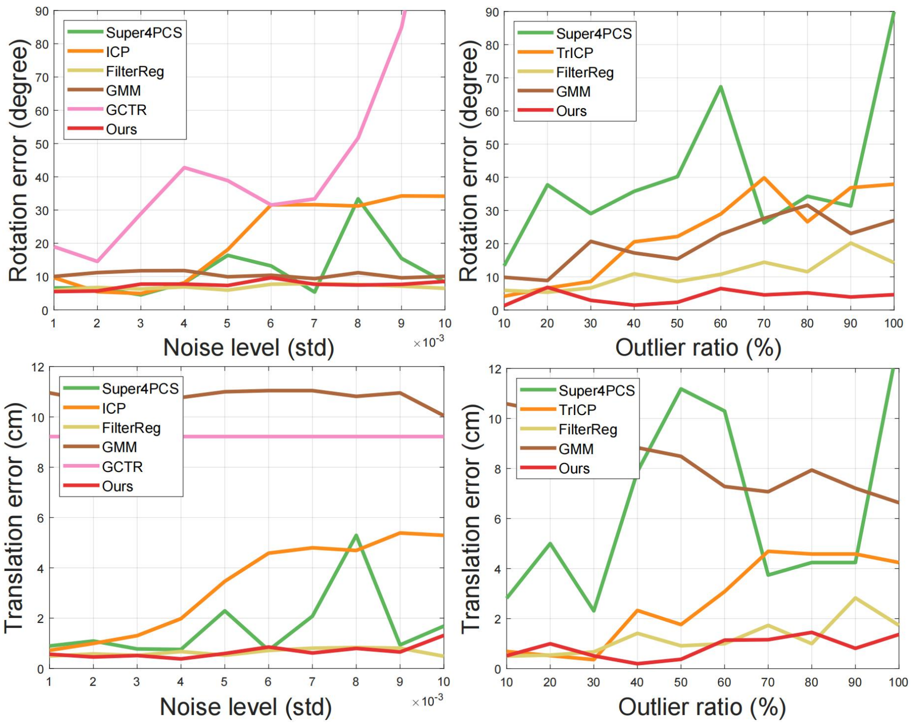
METHOD - RELATION WITH ICP

Lemma 1. *Let* $I = \{1, \cdots, z\} \subseteq Z^+$ *is an index set,* $a_{i \in I} \in \mathbb{R}^+$ *and* $p \in (0, +\infty)$*, then*

$$z^{-\frac{1}{p}} \min_{i \in I} a_i \leq \left(\sum_{i=1}^{z} a_i^{-p}\right)^{-\frac{1}{p}} \leq \min_{i \in I} a_i .$$

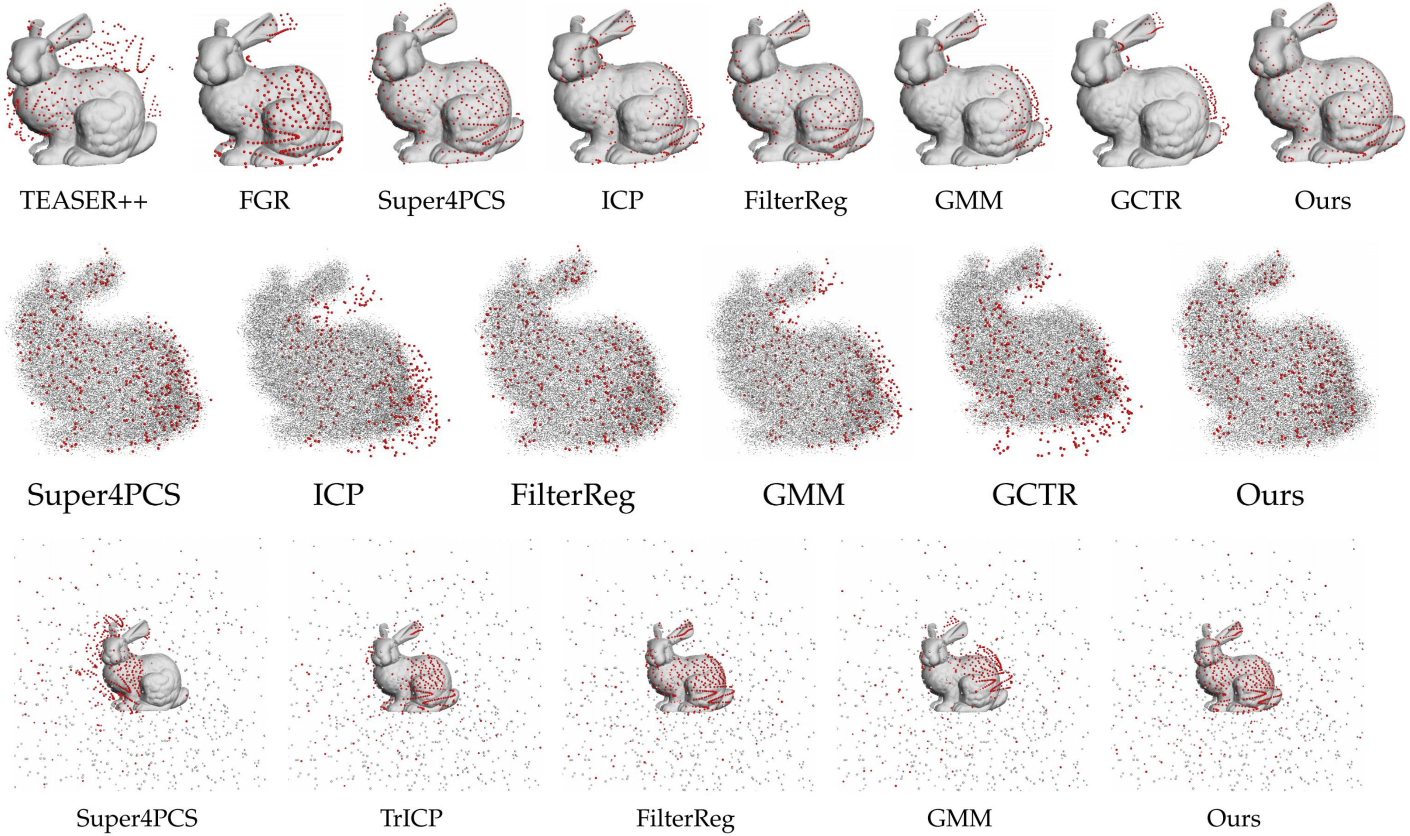
Theorem 1. *The iterative closest point (ICP) algorithm is a special case of the proposed fuzzy registration method when the fuzzier* $m \in (1, +\infty)$ *converges to 1 and* $\mathbf{\Sigma} = \mathbf{I}$.

RESULTS - SYNTHETIC TEST



noise & outlier test

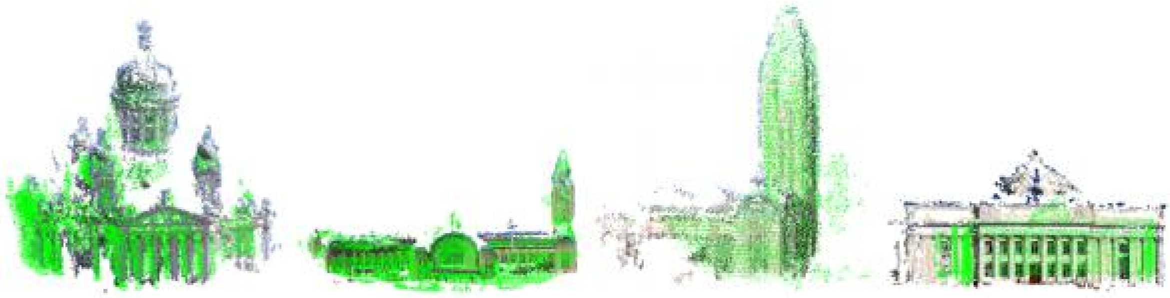
RESULTS - SYNTHETIC TEST



RESULTS - LiDAR & MVS

Real-world Helsinki outdoor dataset [Peng et al. 2014]

Dataset	Metric	Super4PC [22]	TrICP [17]	FilterReg [29]	GMM [26]	GCTR [9]	Ours
Library	RE	165.1815	84.7910	60.2347	4.8023	163.2037	**3.2664**
	TE	0.4297	0.1828	0.2316	0.2039	0.4136	**0.1063**
Cathedral	RE	79.9109	11.6182	**2.3517**	11.8433	121.6016	6.0415
	TE	1.0391	0.2972	**0.1154**	1.4506	1.0503	0.2782
East Station	RE	85.4812	15.2125	1.4468	2.5072	77.6636	**0.9823**
	TE	1.7305	0.1519	0.0409	0.3290	0.2513	**0.0341**
South Station	RE	147.1696	1.9297	0.9860	4.9186	106.8087	**0.8726**
	TE	0.0921	**0.0015**	0.0347	0.0563	0.0489	0.0456



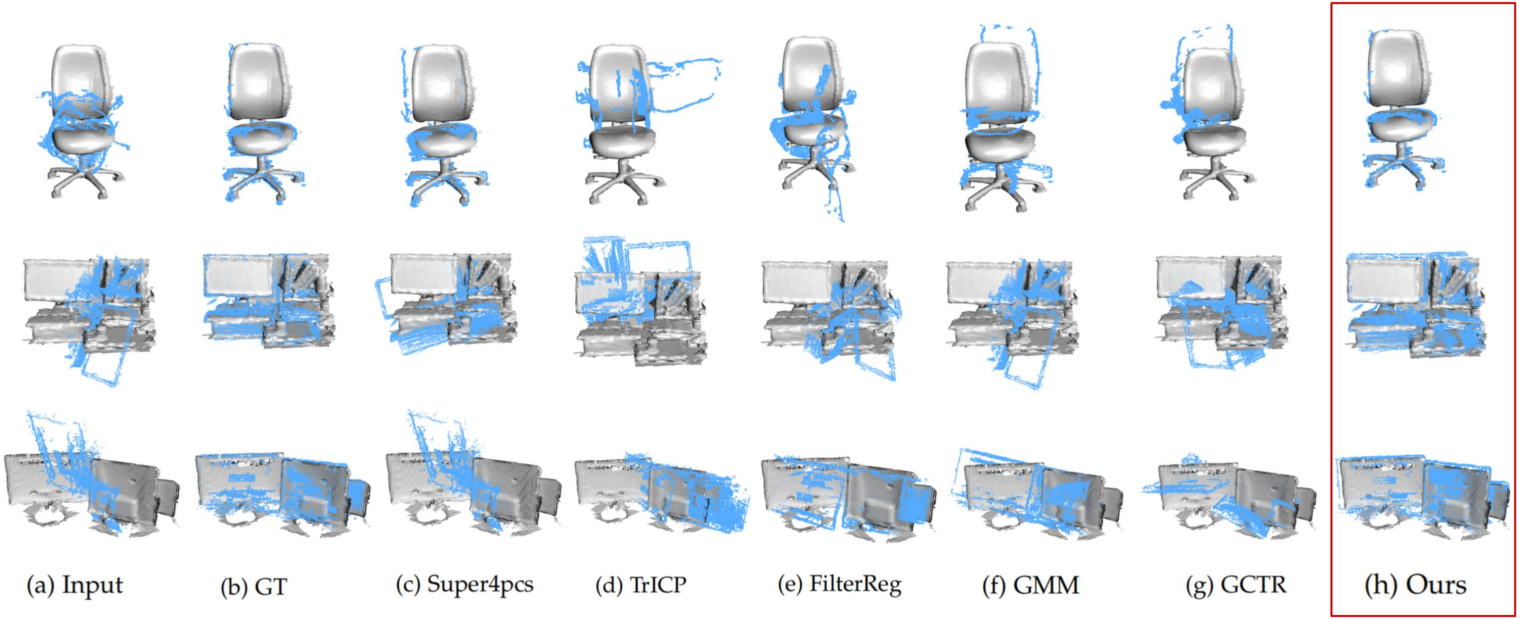
Align real-world LiDAR data to MVS data

RESULTS - LiDAR & KINECT

Real-world 3DCSR indoor dataset [Huang et al. 2019]

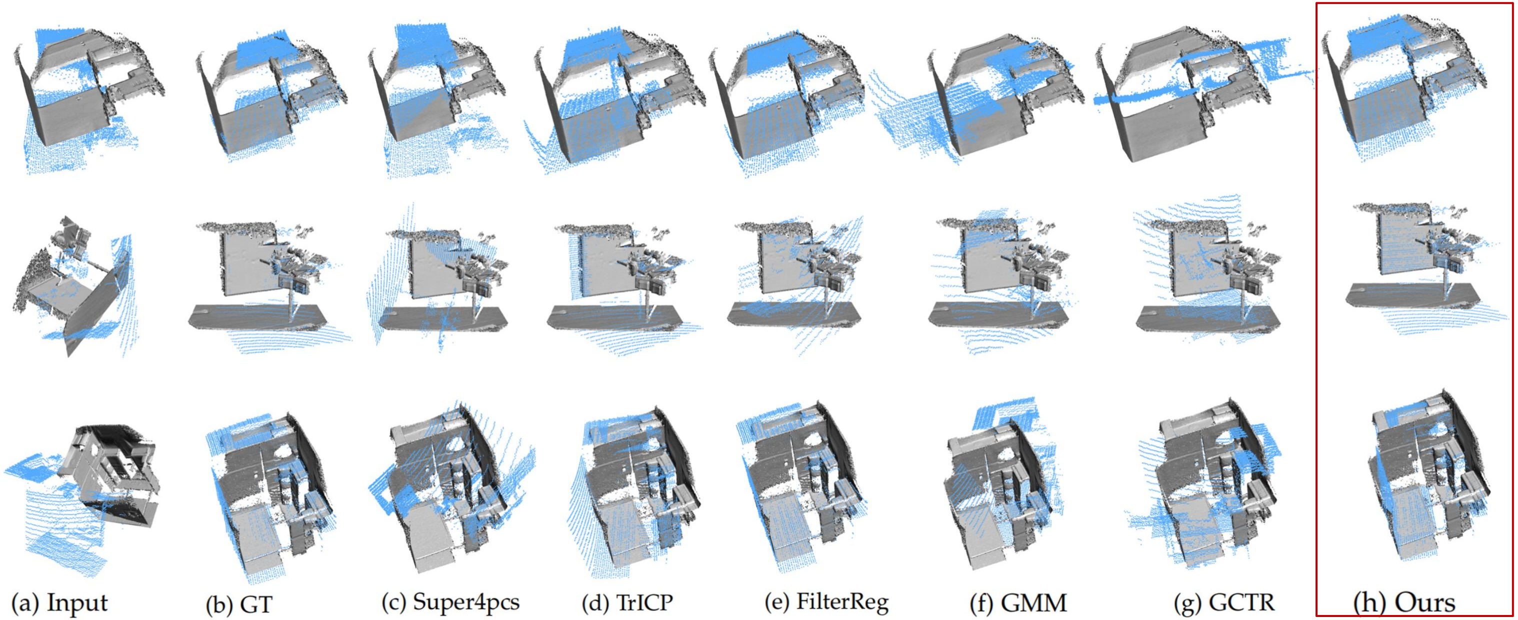
Method	Recall↑	TE↓	RE↓	Time(s)↓
DGR [39]	36.60	**0.04**	4.26	0.87
FMR [58]	17.80	0.10	4.66	0.28
PointNetLK [37]	0.05	0.09	12.54	2.25
FGR [19]	1.49	0.07	10.74	2.23
ESF-64+ICP [43]	24.30	**0.04**	5.71	0.19
JRMPC [59]	0.00	-	-	18.10
RANSAC [10]	3.47	0.13	8.30	**0.03**
GCTR [9]	0.50	0.17	7.46	15.80
Super4PCS [23]	6.93	0.24	6.38	1.70
TrICP [18]	7.92	0.18	6.40	1.26
FilterReg [32]	30.96	0.10	2.45	7.06
GMM [29]	9.41	0.18	7.92	13.34
CICP [45]	2.48	0.28	8.28	0.61
PICP [26]	4.45	0.29	10.85	12.58
Ours	**40.59**	0.06	**2.21**	4.47

RESULTS - LiDAR & KINECT



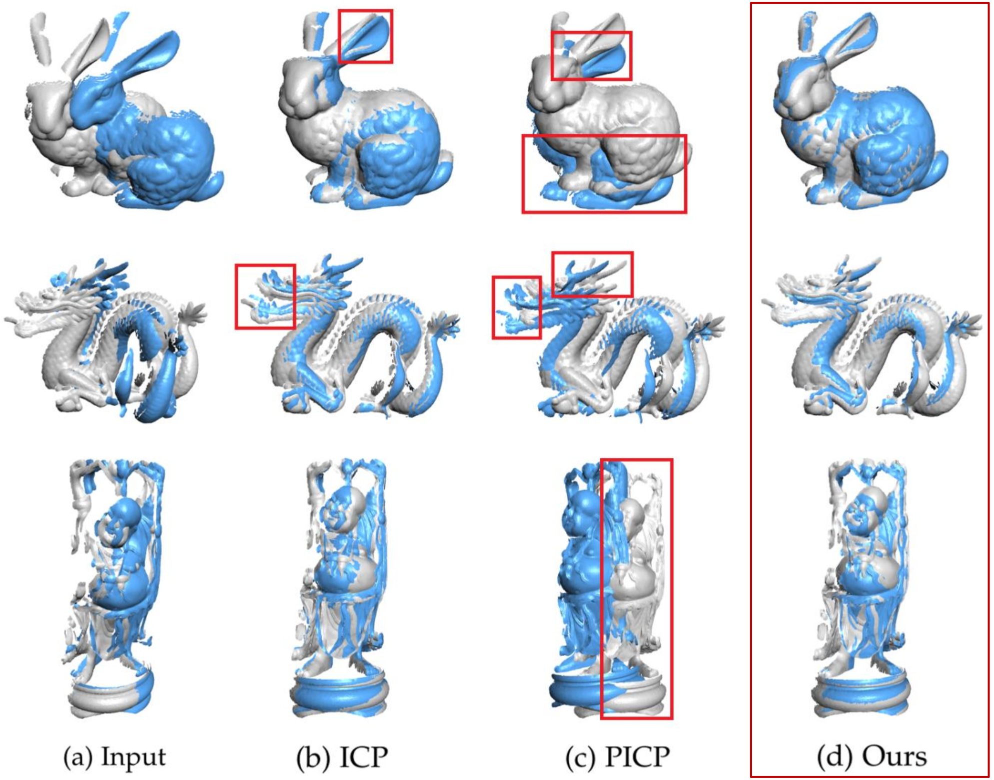
Real-world 3DCSR indoor dataset [Huang et al. 2019]

RESULTS - LiDAR & KINECT



Real-world 3DCSR indoor dataset [Huang et al. 2019]

RESULTS - SAME MODALITY

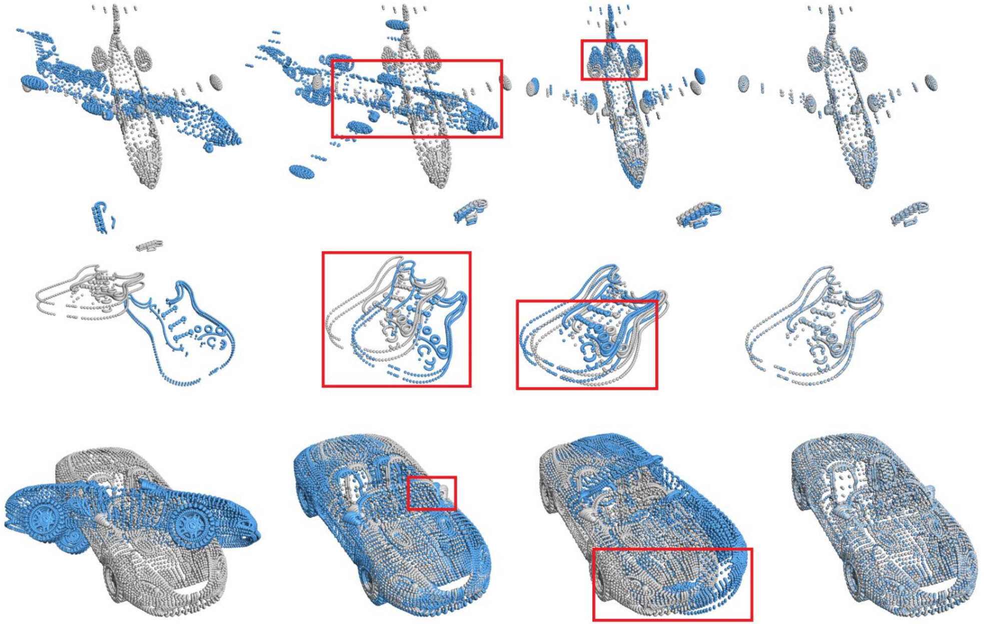


Dataset / Method	Bunny		Dragon		Happy Buddaha	
	RE↓	TE↓	RE↓	TE↓	RE↓	TE↓
ICP [13]	4.2717	0.0022	4.6766	**0.0060**	5.0297	**0.0016**
PICP [26]	21.5870	0.1286	17.7415	0.0184	25.7583	0.0154
Ours	**1.2439**	**0.0013**	**1.3350**	0.0104	**0.4647**	0.0020

Dataset / Method	Airplane		Guitar		Car	
	RE↓	TE↓	RE↓	TE↓	RE↓	TE↓
ICP [13]	54.3791	0.5157	19.2938	0.1412	5.5450	0.0762
PICP [26]	10.8500	0.0844	5.2585	0.0429	193.3783	0.2240
Ours	**1.0001e-6**	**1.5128e-7**	**2.3286e-6**	**9.5704e-6**	**2.3454e-6**	**8.9725e-8**

The Stanford 3D scanning repository

RESULTS - SAME MODALITY

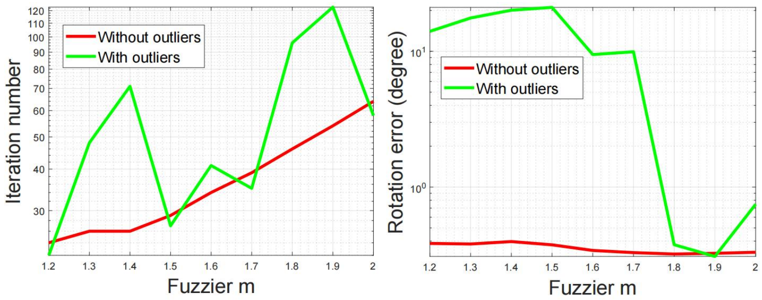


(a) Input (b) ICP (c) PICP (d) Ours

ModelNet 40

RESULTS - INVESTIGATION OF THE FUZZIER

$$\min_{\phi_{ra},\sigma^2} J(\phi_{ra},\sigma^2) = \sum_{i=1}^{M}\sum_{j=1}^{N} u_{ij}^m \|\phi_{ra}(\boldsymbol{x}_i) - \boldsymbol{y}_j\|_{\boldsymbol{\Sigma}_j}^2 + \beta \log|\boldsymbol{\Sigma}_j|$$



1. As m increases, iterations increase, rotation error decreases

2. Recommend $m \in [1.8, 2]$ in use

LIMITATIONS & FUTURE WORK

- Boost the efficiency with parallel optimization
- Compile more cross source data
- Extend to non-rigid scenarios using unsupervised clustering

CONCLUSION

- Take structure similarity handling cross-modality registration
- Formulate cross-modality registration (coarse&fine) as a clustering process
- Demonstrate impressive performance on common data modalities

THANK YOU FOR YOUR ATTENTION!