Multi-Body Neural Scene Flow

Kavisha Vidanapathirana[1,2,], Shin-Fang Ch'ng[3], Xueqian Li[3], Simon Lucey[3]*

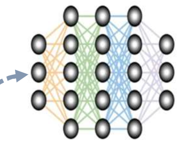
[1]Queensland University of Technology, [2]CSIRO Robotics, [3]The University of Adelaide

*Work done during an internship at The Australian Institute for Machine Learning, The University of Adelaide.

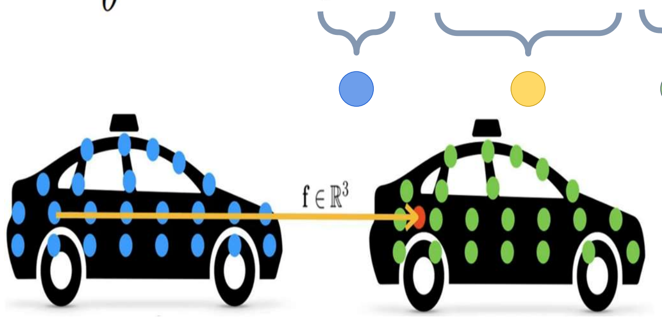
3DV 2024

Background: Neural Scene Flow Prior

"Spatially smooth" scene flow via the smoothness bias of the coordinate MLP



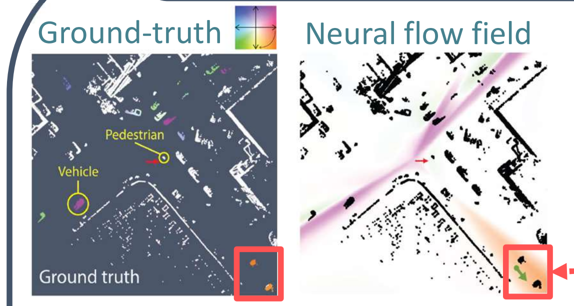
$$\theta^* = \underset{\theta}{\operatorname{argmin}}\ \mathcal{L}_{CD}\Big(\mathbf{P}_1 + \Phi(\mathbf{P}_1;\theta), \mathbf{P}_2\Big)$$



Can we enforce multi-body rigidity in neural scene flow without:

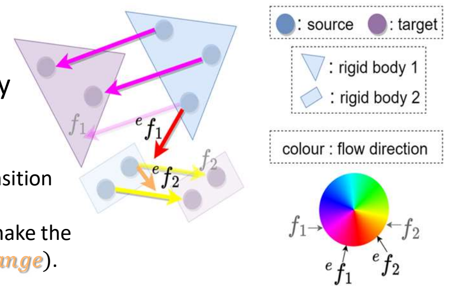
- Sacrificing the continuous flow field?
- Constraining the *SE(3)* parameters of each rigid body?
- Accurately estimating the rigid bodies?

Motivation: Rigidity in Neural Scene Flow



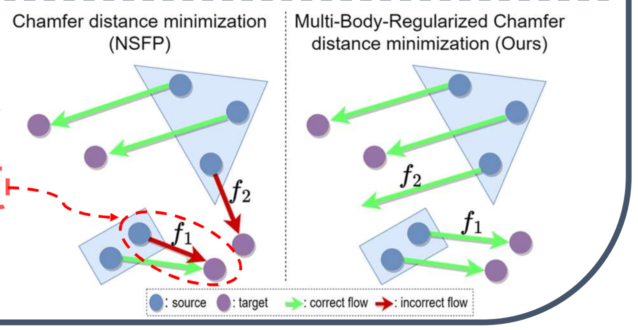
"Spatial smoothness" prior is unaware of object boundaries

Enforcing smoothness at object boundaries may lead to physically implausible flow predictions



- Ground-truth flow contains a sharp transition from f_1 (*purple*) to f_2 (*yellow*).
- The neural prior tends to (incorrectly) make the transition smooth - ${}^e f_1$ (*red*) to ${}^e f_2$ (*orange*).

Chamfer distance optimization may violate multi-body rigidity due to nearest neighbor correspondence assignment

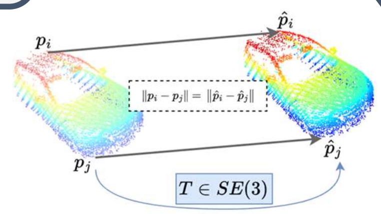


Approach: Rigid flow via isometry

Key insight:

The flow of rigid bodies should maintain an isometry

$$\|\mathbf{p}_i - \mathbf{p}_j\| = \|\hat{\mathbf{p}}_i - \hat{\mathbf{p}}_j\|\ \forall i,j \implies \hat{\mathbf{p}}_i = \mathbf{T} \circ \mathbf{p}_i\ \ \forall i.$$



Encouraging isometry in the flow predictions for rigid bodies will ensure *SE(3)* rigidity without having to estimate/constrain the *SE(3)* parameters

Formulation:

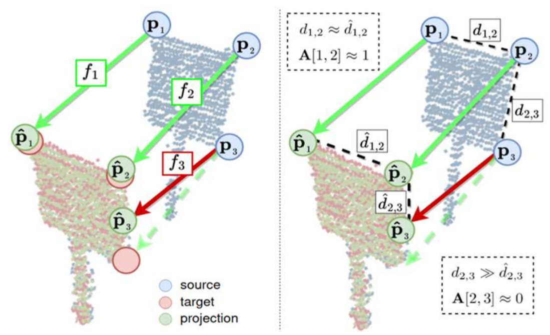
For each rigid body, score *s(**C**,**F**)* represents the degree of preservation of isometry for its point cluster **C** and its associated flow prediction **F**.

$$s(\mathbf{C},\mathbf{F}) = \frac{1}{n}\mathbf{v}^{*\intercal}\mathbf{A}\mathbf{v}^*$$

$$\mathbf{A}[i,j] = \left[1 - \frac{(d_{i,j} - \hat{d}_{i,j})^2}{d_{thr}^2}\right]_+$$

*s(**C**,**F**)* = 1 ⇒ exact isometry

0 < *s(**C**,**F**)* < 1 ⇒ approximate isometry



Final objective:

Maximize the score *s(**C**,**F**)* for all clusters in the scene while reducing Chamfer distance.

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\ \mathcal{L}_{CD}\Big(\mathbf{P}_1 + \Phi(\mathbf{P}_1;\theta), \mathbf{P}_2\Big) + \omega \cdot \mathcal{L}_{MB}\Big(\mathbf{P}_1;\theta\Big)$$

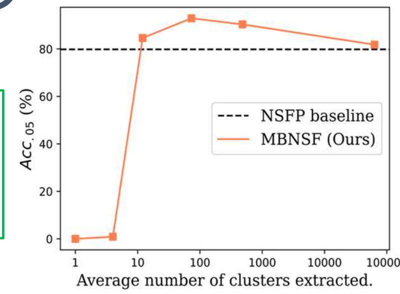
where

$$\mathcal{L}_{MB}(\mathbf{P}_1;\theta) = -\log\big(s_{avg}(\mathbf{P}_1;\theta)\big)$$

$$s_{avg}(\mathbf{P}_1;\theta) = \frac{1}{m}\sum_{i=1}^{m} s(\mathbf{C}_i,\mathbf{F}_i)$$

Results Quantitative

Robust to over-clustering ⇒ Accurate rigid body estimation is not needed



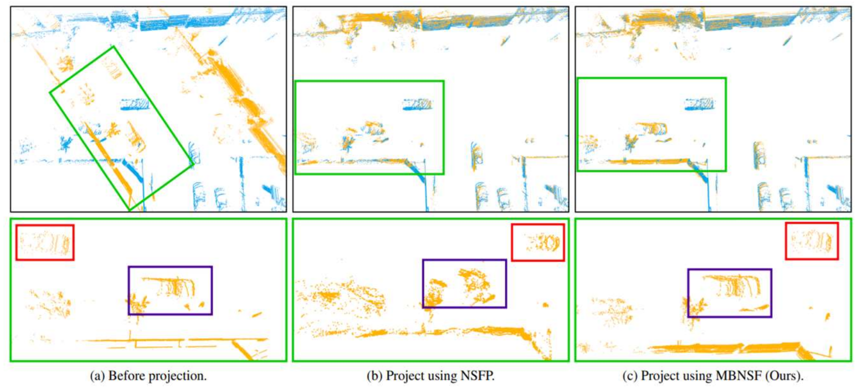
Performance improvement on any dataset, on any number of points

Dataset	Method	Supervision	Num. Points	EPE [m] ↓	$Acc_{.05}$ [%] ↑	$Acc_{.10}$ [%] ↑	θ_ϵ[°] ↓
Waymo	FLOT [37]	Full	8192	0.702	2.46	11.30	0.808
	PointPWC-Net [54]	Self	8192	4.109	0.05	0.36	1.742
	WsRSF [16]	Weak	8192	0.414	35.47	44.96	0.527
	NSFP [26]	None	8192	0.098	69.34	85.95	0.302
	MBNSF (Ours)	None	8192	**0.097**	**74.15**	**89.24**	**0.295**
	NSFP [26]	None	All	0.091	76.61	89.00	0.282
	MBNSF (Ours)	None	All	**0.066**	**82.29**	**92.44**	**0.277**
Argoverse	FLOT [37]	Full	8192	0.796	2.30	9.87	0.929
	PointPWC-Net [54]	Self	8192	5.724	0.02	0.15	1.147
	WsRSF [16]	Weak	8192	0.416	34.52	43.10	0.558
	NSFP [26]	None	8192	0.056	70.30	89.57	0.265
	MBNSF (Ours)	None	8192	**0.051**	**79.36**	**92.37**	**0.264**
	NSFP [26]	None	All	0.090	64.97	83.09	0.230
	MBNSF (Ours)	None	All	**0.033**	**89.34**	**95.91**	**0.168**

Supports both pairwise scene flow and direct 4D trajectory estimation

Approach	Method	$Acc_{.5}$ [%] ↑	$Acc_{1.0}$ [%] ↑
Scene flow integration (forward-Euler)	NSFP [47]*	45.28	59.52
	NSFP†	38.09	54.20
	MBNSF (Ours)	**61.11**	**74.34**
Direct 4D trajectory prediction	NTP [47]*	52.28	69.88
	NTP†	48.06	62.45
	MBNT (Ours)	**59.82**	**73.57**

Results Qualitative



(a) Before projection. (b) Project using NSFP. (c) Project using MBNSF (Ours).

Preserves multi-body rigidity, even across long sequences.

Implementation

Our code is publicly available: