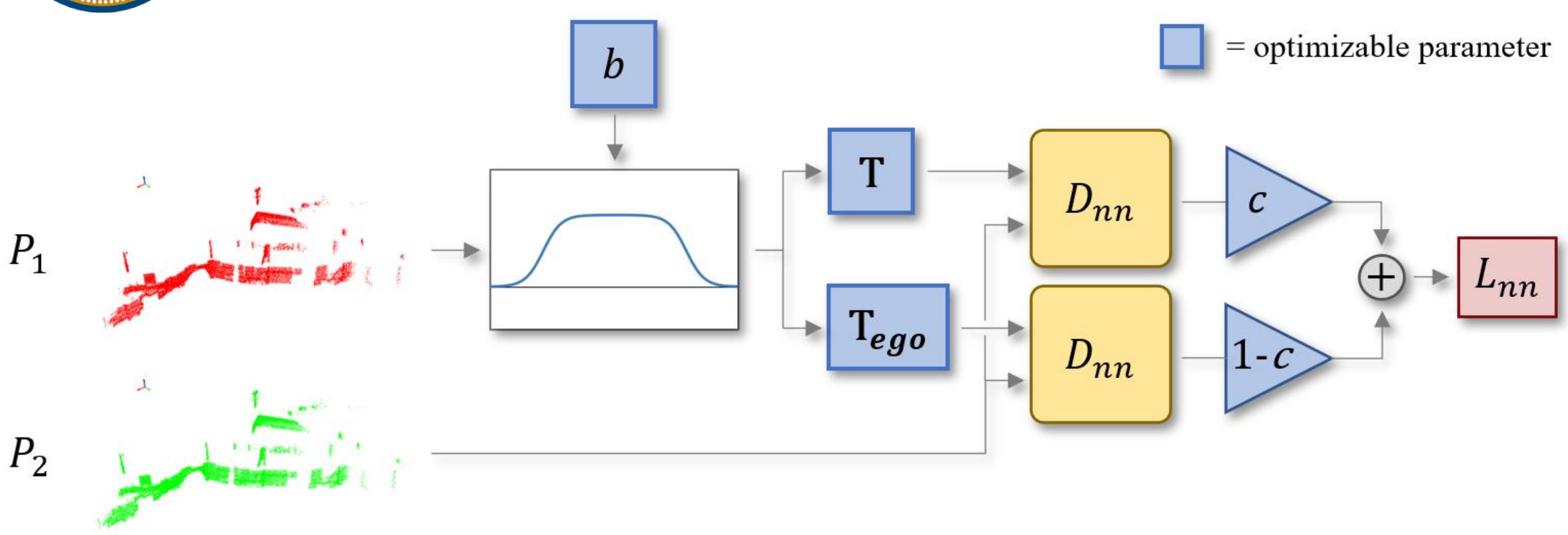
RSF: Optimizing Rigid Scene Flow From 3D Point Clouds Without Labels

David Deng and Avideh Zakhor
University of California, Berkeley



Scene Flow Parameters

$$\mathbf{T}_{ego} \in SE(3), \mathbf{T}_i \in SE(3)$$

$$\mathbf{b}_i = \{c_i, x_i, y_i, z_i, w_i, l_i, h_i, \theta_i\}$$

$$\text{for } i = 1 \ldots k$$

$c_i =$ confidence, $\{x_i, y_i, z_i\} =$ position, $\{w_i, l_i, h_i\} =$dimensions, $\theta_i =$heading angle, $k =$number of boxes

Overview of our loss function for a single bounding box. Terms in blue blocks are optimizable scene flow parameters. b: bounding box parameters, $\mathbf{T}$: bounding box's rigid transformation, $\mathbf{T}_{ego}$: ego-motion transformation, c: box's confidence score, and the plot refers to our differentiable bounding box approximation. From P_1, we differentiably select the points inside the bounding box and transform them using $\mathbf{T}$ and $\mathbf{T}_{ego}$. Then we compute the nearest neighbor distance (NND) between the two transformed point sets and P_2. Lastly, we weigh the two NNDs by c and $1-c$ respectively and sum them to compute the loss. Our total loss is the sum of per box losses in addition to shape, heading, angle, and mass auxiliary losses.

Problem

Given a pair of 3D point clouds, predict the scene flow between them without any labels.

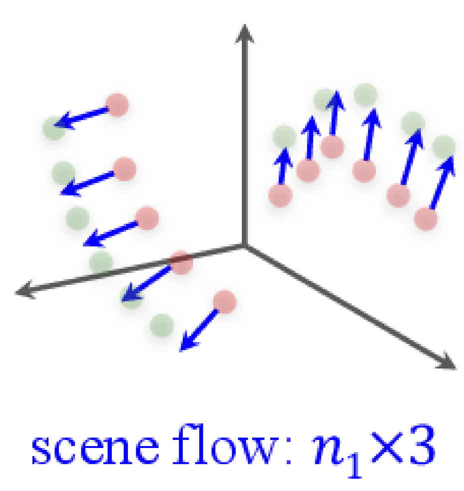
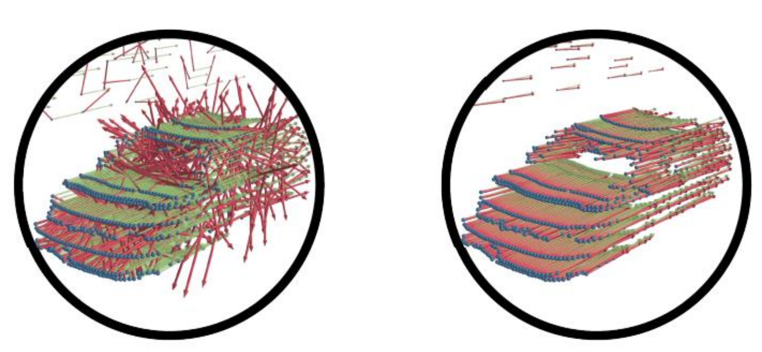


Image taken from [1].

Background

- Self-supervised scene flow from point clouds
  - Minimize NND over pointwise motion vectors
  - Smoothness and geometry regularization, cycle consistency
- Optimization based approaches
  - Neural Scene Flow Prior [2] trains a DNN for each scene, using it as an implicit regularizer



(a) Without the graph Laplacian (b) With the graph Laplacian

Image taken from [3]

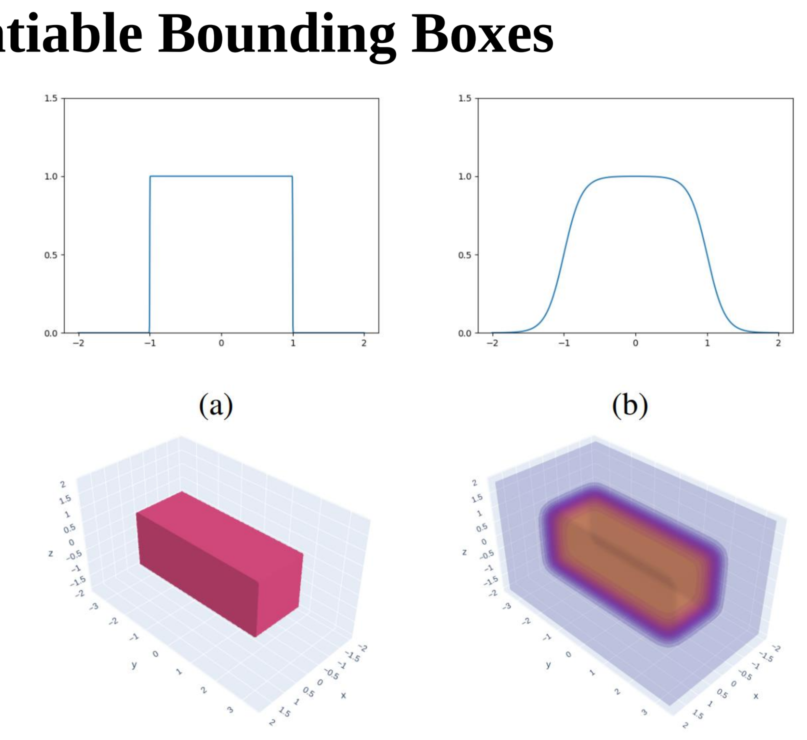
Approach

Observation: scenes are comprised of independently moving rigid objects

Instead of pointwise vectors, we parameterize scene flow at the object level and parameterize objects as bounding boxes

- Lower dimension optimization space
- Constrains scene flow to be physically coherent
- Solve for parameters by optimizing them over the NND with gradient methods

Differentiable Bounding Boxes

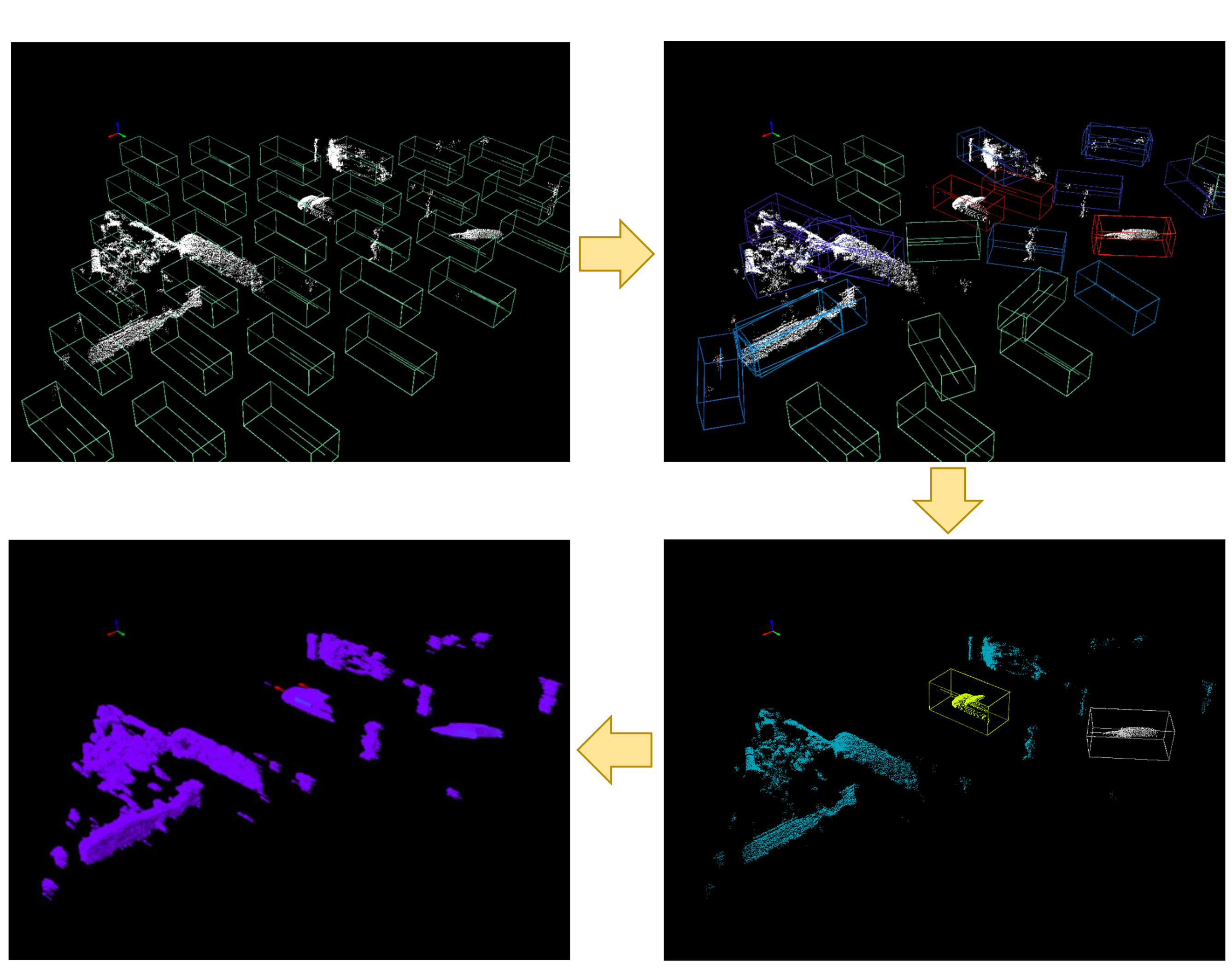


Visualizations of non-differentiable vs differentiable bounding boxes in 1 and 3 dimensions: (a) non-differentiable 1D bounding line; (b) differentiable 1D bounding line; (c) non-differentiable 3D bounding box; (d) differentiable 3D bounding box.

Inference

Post processing: prune empty boxes, non-maximum suppression, assign points to objects

Scene flow: $\mathbf{f}_i = \mathbf{R}_i \mathbf{p}_i + \mathbf{t}_i - \mathbf{p}_i$



Results

Table 1: Scene flow evaluation.

Dataset	Method	Supervision/ Approach	Training Data	EPE3D ↓	Acc3DS ↑	Acc3DR ↑	Outliers ↓
StereoKITTI	FlowNet3D [27]	Full	FT3D	0.177	0.374	0.668	0.527
	HPLFlowNet [13]	Full	FT3D	0.117	0.478	0.778	0.410
	PointPWCNet [50]	Full	FT3D	0.069	0.728	0.888	0.265
	FLOT [36]	Full	FT3D	0.056	0.755	0.908	0.242
	EgoFlow [54]	Full	FT3D	0.069	0.670	0.879	0.404
	FlowStep3D [51]	Full	FT3D	0.055	0.805	0.925	0.149
	HCRF-Flow [39]	Full	FT3D	0.053	0.863	0.944	0.180
	WeaklyRigidFlow [12]	Full	FT3D	0.042	0.849	0.959	0.208
	PointPWCNet [50]	Self	FT3D	0.255	0.238	0.496	0.686
	EgoFlow [54]	Self	FT3D	0.415	0.221	0.372	0.810
	FlowStep3D [51]	Self	FT3D	0.102	0.708	0.839	0.246
	SLIM [1]	Self	RawKITTI	0.121	0.518	0.796	0.402
	SLIM*[1]	Self	RawKITTI	0.067	0.77	0.934	0.249
	RigidFlow [24]	Self	FT3D	0.062	0.724	0.892	0.262
	Chamfer*	Optimization	-	0.991	0.056	0.071	0.942
	PointPWCNet [50]	Optimization	-	0.657	0.357	0.405	0.72
	NSFP [25]	Optimization	-	0.036	0.912	0.961	0.154
	NSFP*[25]	Optimization	-	0.034	0.914	0.962	0.151
	Ours	Optimization	-	0.035	0.932	0.971	0.146
	Ours*	Optimization	-	**0.017**	**0.973**	**0.989**	**0.096**
LidarKITTI	PointPWCNet [50]	Full	FT3D	0.390	0.387	0.550	0.653
	FLOT [36]	Full	FT3D	0.653	0.155	0.313	0.837
	WeaklyRigidFlow [12]	Weak	SemKITTI	0.094	0.784	0.885	0.314
	ExploitingRigidity [8]	Weak	SemKITTI	**0.071**	0.824	0.913	0.295
	Chamfer*	Optimization	-	0.944	0.022	0.057	0.992
	PointPWCNet [50]	Optimization	-	0.734	0.248	0.347	0.845
	NSFP*[25]	Optimization	-	0.142	0.688	0.826	0.385
	Ours*	Optimization	-	0.085	**0.883**	**0.929**	**0.239**
nuScenes	Chamfer*	Optimization	-	0.879	0.035	0.082	0.976
	PointPWCNet*[50]	Optimization	-	0.615	0.199	0.328	0.86
	NSFP *[25]	Optimization	-	0.177	0.374	0.668	0.527
	Ours *	Optimization	-	**0.107**	**0.717**	**0.862**	**0.321**

* methods that use the entire point cloud. All other methods downsample to 8,192 points.

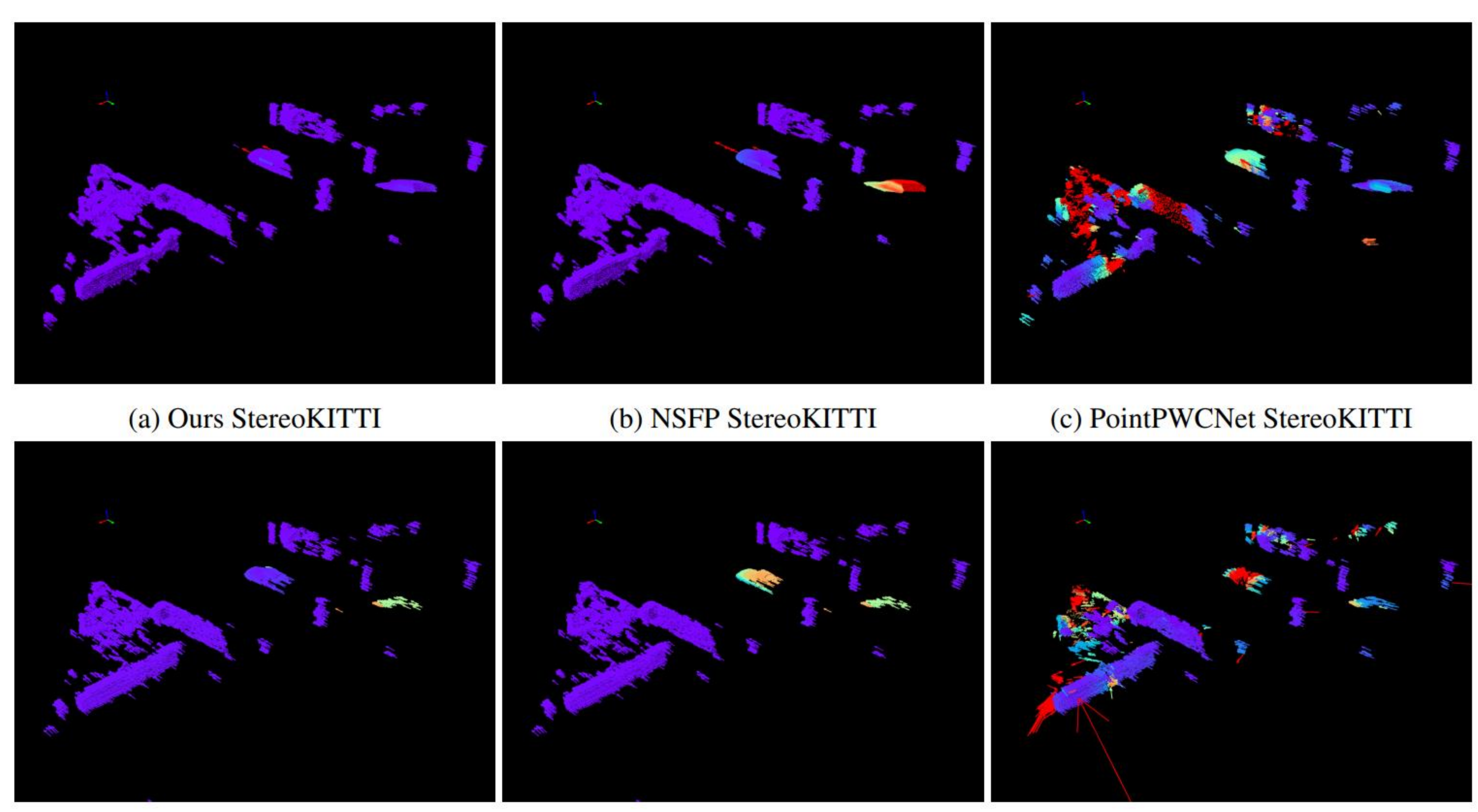
Table 2: Motion segmentation results on StereoKITTI.

Method	mIoU ↑	Accuracy ↑
SLIM [1]	42.9	60.1
Ours	**86.6**	**92.9**

Table 3: Ego-Motion Estimation Evaluation on SemanticKITTI.

Method	Rotation Error (○) ↓	Translation Error (m) ↓	Rotation Accuracy ↑	Translation Accuracy ↑
ICP [3]	0.244	0.122	0.906	0.878
Ours	**0.235**	**0.107**	**0.916**	**0.94**

Visualizations



(a) Ours StereoKITTI (b) NSFP StereoKITTI (c) PointPWCNet StereoKITTI

(d) Ours LidarKITTI (e) NSFP LidarKITTI (f) PointPWCNet LidarKITTI

(g) Error Colorbar

Visualization of scene flow predictions for our method, NSFP, and the PointPWCNet loss function under direct optimization on a scene in KITTI. Color indicates the EPE3D of the prediction, with red indicating high error and purple indicating low error. For StereoKITTI, the colorscale ranges from 0-0.5 m error, while for LidarKITTI, it ranges from 0-1 m.

Conclusion

We propose a novel method for optimizing object-level rigid scene flow without labels. Our method achieves state-of-the-art accuracy by constraining the scene flow to be physically consistent, and it simultaneously detects moving objects and computes ego-motion without any labels.

References

[1] Xingyu Liu, Charles R Qi, and Leonidas J Guibas. Flownet3d: Learning scene flow in 3d point clouds. In *In Computer Vision and Pattern Recognition (CVPR)*, 2019.
[2] Xueqian Li, Jhony Kaesemodel Pontes, and Simon Lucey. Neural scene flow prior. *Advances in Neural Information Processing Systems*, 34, 2021.
[3] Jhony Kaesemodel Pontes, James Hays, and Simon Lucey. Scene flow from point clouds with or without learning. *International Conference on 3D Vision (3DV)*, 2020.