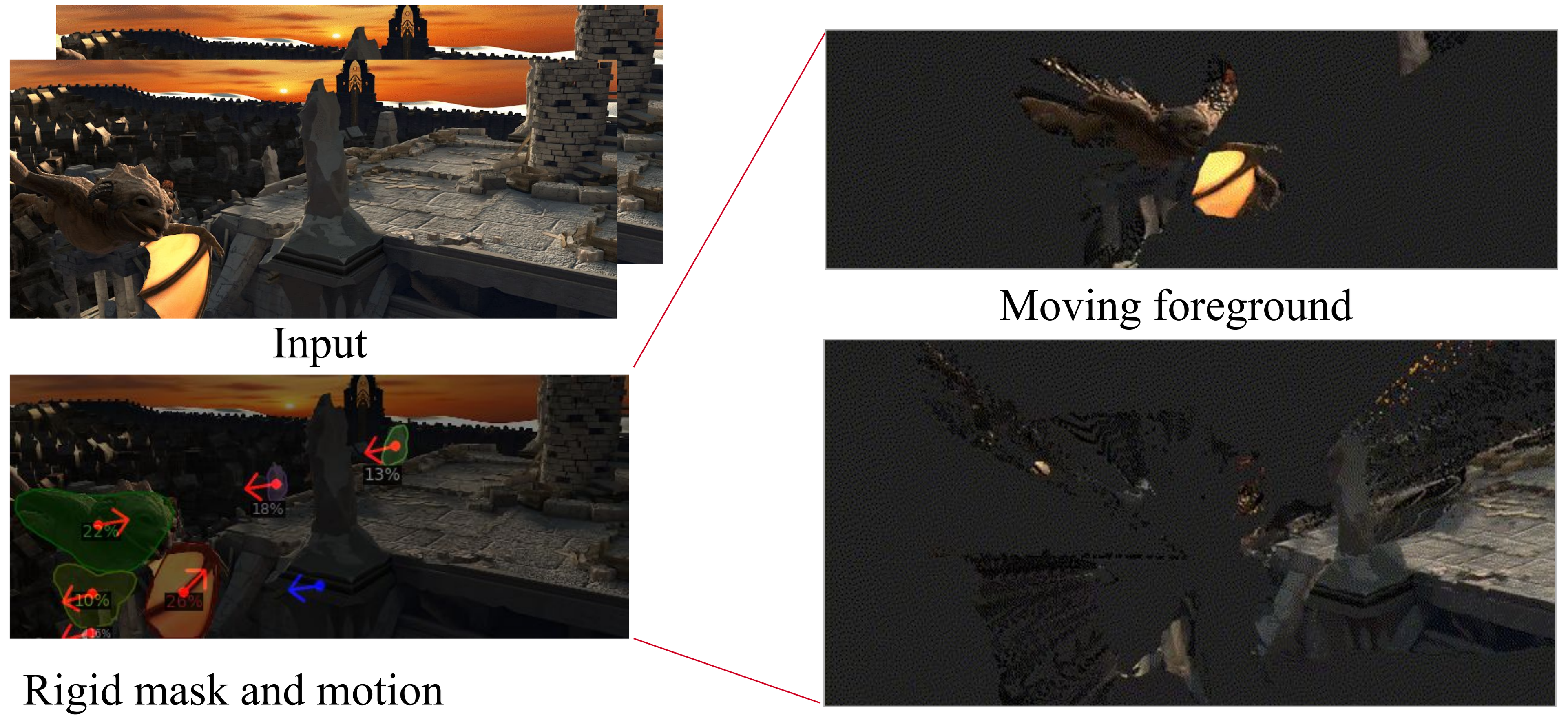
Learning to Segment Rigid Motions from Two Frames

Gengshan Yang and Deva Ramanan

Code available at : https://github.com/gengshan-y/rigidmask

Goal: Scene decomposition based on two-frame rigidity



Input

Moving foreground

Rigid mask and motion
(blue: ego-motion; red: object motion)

Static background reconstruction

Related work

Geometric Motion segmentation (2D motion angle) [1,2]

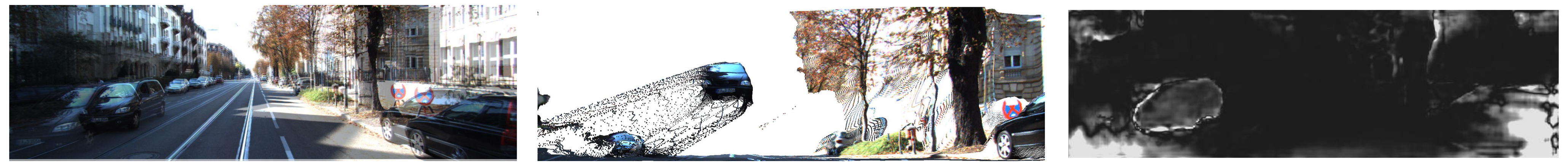
Appearance-based detectors (fires on objects ***able to move***)

Ours (fires only on ***moving*** objects)

[1] P Bideau and E Learned-Miller. "It's moving! A probabilistic model for causal motion segmentation in moving camera videos." ECCV. 2016.
[2] J Wulff, L Sevilla-Lara, MJ Black. "Optical flow in mostly rigid scenes." CVPR. 2017.

Challenges

- Degeneracies in geometric motion segmentation.
  - Epipolar constraints fail when translation is close to zero.
  - Points moving along epipolar lines cannot be distinguished.
- Noisy motion correspondences and camera egomotion estimates.



Two frame overlay

Flow-triangulated reconstruction

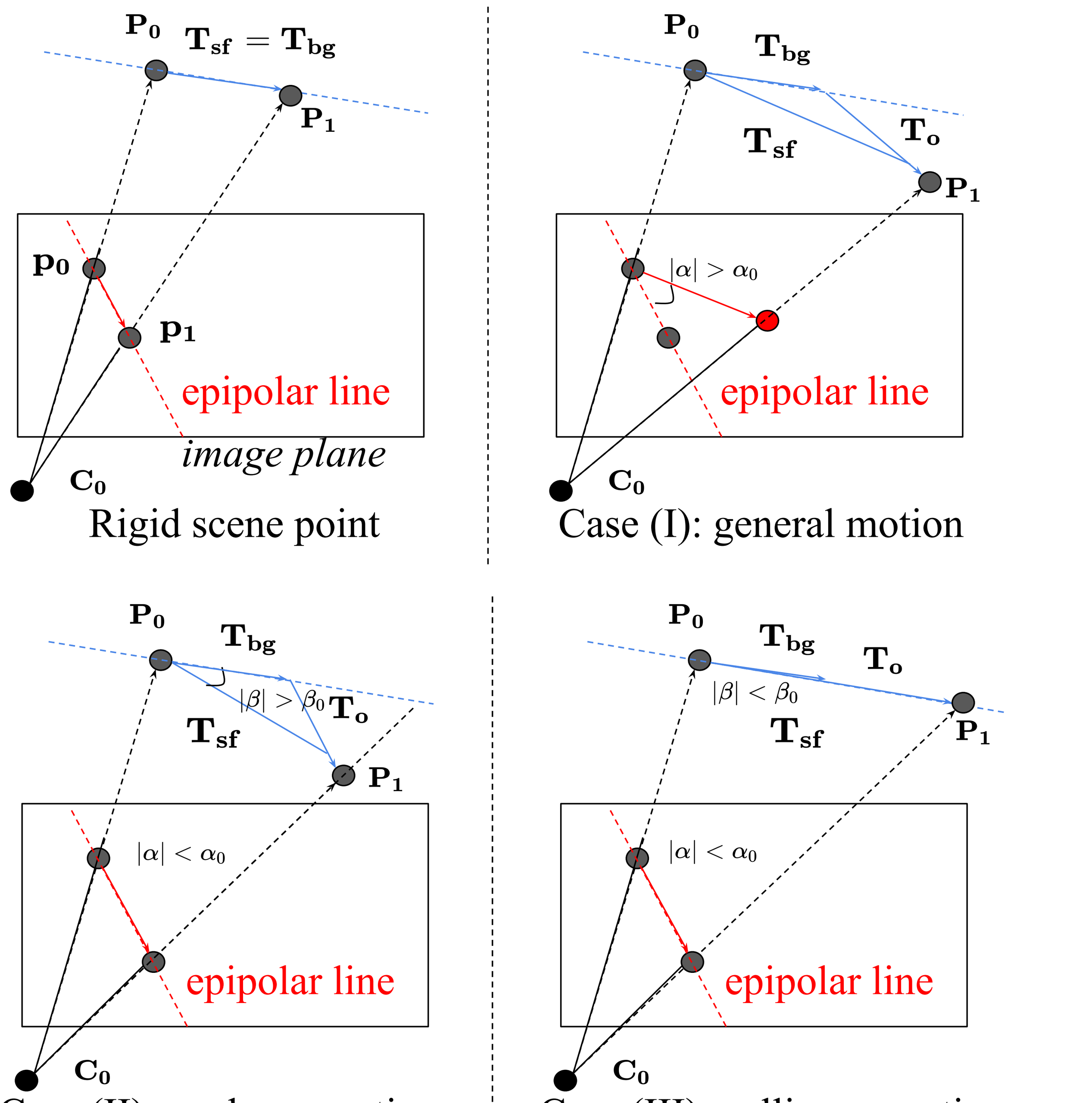
Sampson err. ($\mathbf{p1}$,$\mathbf{p2}$;$\mathbf{F}$)
(white → moving)

Analysis

When can a moving scene point P be identified from a moving camera?

$\mathbf{T_{bg}} := -\mathbf{T_c}$ (bg motion due to camera motion)

$\mathbf{T_{sf}} := \mathbf{P_1} - \mathbf{P_0}$ (3D scene flow assuming no cam rotation)



Rigid scene point

Case (I): general motion

Case (II): coplanar motion

Case (III): collinear motion

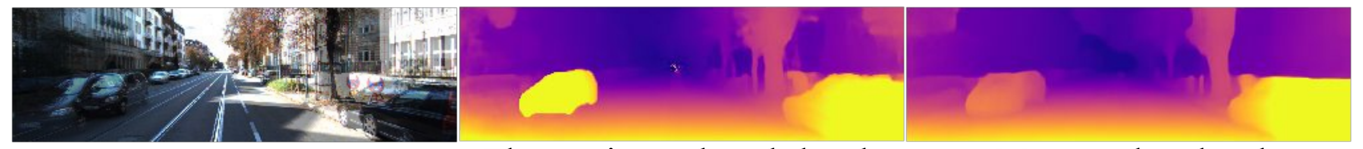
Degeneracies and rigidity cost maps

- **Small translation:** Rotational homography $c_H(\mathbf{p_0}, \mathbf{p_1}; \mathbf{H_R})$
- **Coplanar motion** (ambiguous w/o relative depth τ=Z1/Z0)

Two frame overlay　Sampson error　3D angle costs

$$c_{\text{3D-angle}}(\mathbf{p_0}, \mathbf{p_1}, \tau; \tilde{\mathbf{T}}_{\mathbf{bg}}, \mathbf{H_R}) = |\angle(\tilde{\mathbf{T}}_{\mathbf{bg}}, \tilde{\mathbf{T}}_{\mathbf{sf}})|, \quad \tilde{\mathbf{T}}_{\mathbf{sf}} = \mathbf{K_0}^{-1}(\tau \mathbf{H_R} \mathbf{p_1} - \mathbf{p_0})$$

- **Collinear motion** (ambiguous w/o 1st frame depth Z0)

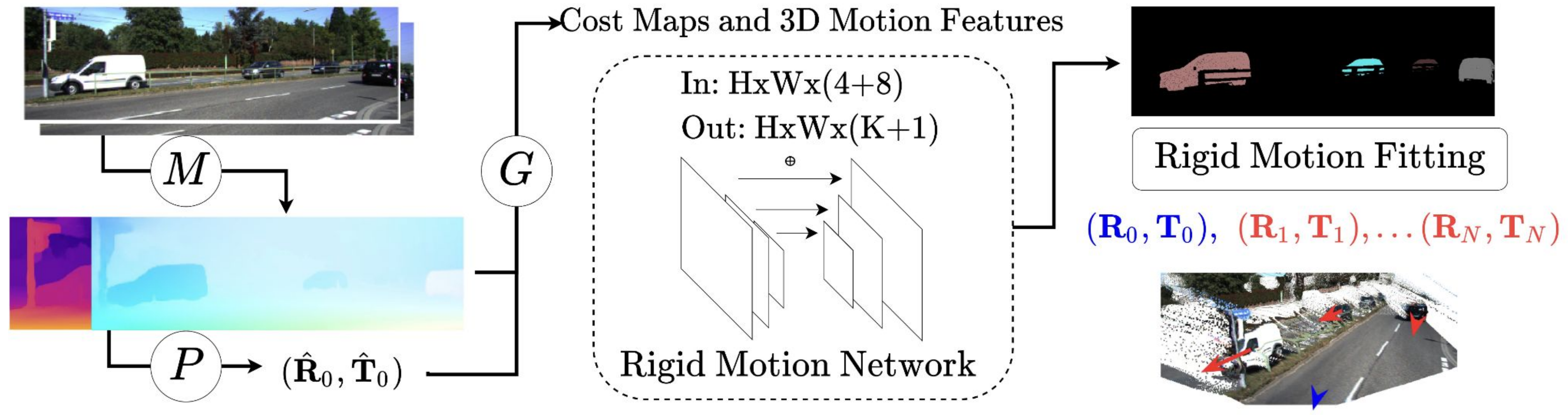


Two frame overlay　Flow-triangulated depth　Monocular depth

$$c_{\text{depth}}(\mathbf{p_0}, \mathbf{p_1}, Z_0; \tilde{\mathbf{T}}_{\mathbf{bg}}, \mathbf{R_{bg}}) = |\log(\frac{Z_0}{\gamma Z_{flow}})|, \quad Z_{flow} = \text{triangulate}(\mathbf{p_0}, \mathbf{p_1}; \tilde{\mathbf{T}}_{bg}, \mathbf{R_{bg}})$$

"scale factor to align two depth maps for solving monocular scale ambiguity

Pipeline



1) Initialization　2) Rigidity Estimation　3) Rigid Body Scene Flow

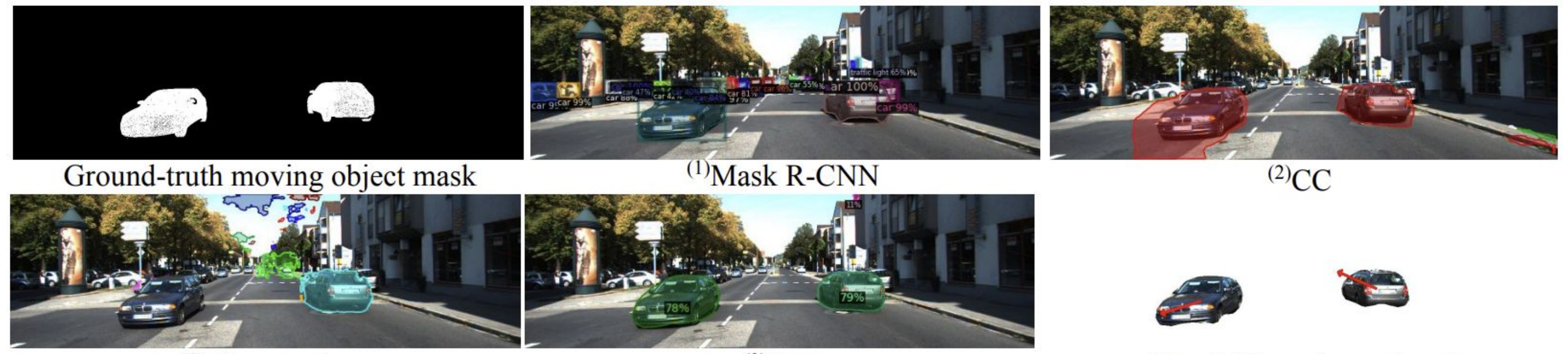
Losses

$$L = L_{\text{binary}} + L_{\text{center}} + L_{\text{polar}}$$

Ground-truth foreground mask　Ground-truth object centers　Ground-truth object contours

Results

Rigid motion segmentation



Ground-truth moving object mask　(1)Mask R-CNN　(2)CC

(3)Flow angle　(3)Ours　Our rigid motion estimation

	Method	K: obj ↑	K: bg ↑	S: bg ↑
(1)	Mask R-CNN [57]	88.20	96.42	81.98
	U^2 (Saliency) [37]	64.80*	93.34	82.01
	MR-Flow-S (K) [58]	75.59*	94.70‡	76.11
	MR-Flow-S (S) [58]	11.11*	84.72	92.64‡
(2)	FSEG [24]	85.08*	96.27	80.22
	MAT-Net [67]	68.40*	93.08	77.95
	COSNet [30]	66.67*	93.03	80.86
	CC [39]	50.87*	85.50	✗
	RTN [31]	34.29*	84.44	64.86
(3)	FSEG-Motion [24]	61.29	89.41	78.25
	CC-Motion [39]	42.99	74.06	✗
	Flow angle [5, 58]	25.83	85.52	74.23
	Ours	**90.71**	**97.05**	**86.72**

Stereo scene flow

Method	D1* ↓	D2 ↓	Fl ↓	SF ↓
PRSM [54]	4.27	6.79	6.68	8.97
OpticalExp [63]	1.81	4.25	6.30	8.12
DRISF [32]	2.55	4.04	4.73	6.31
Ours Mask R-CNN	1.89	3.42	4.26	5.61
Ours Rigid Mask	1.89	**3.23**	**3.50**	**4.89**

Ablation study

Method	K: obj ↑	K: bg ↑	S: bg ↑
Reference	**89.53**	**97.22**	**84.63**
(1)w/o cost maps	88.66	96.59	76.81
(2)w/o uncertainty	85.09	95.72	77.25
(3)w/o monocular depth	84.46	94.84	76.14
(4)w/o expansion (MoA [6])	81.28	95.50	77.00
(5)w/o learning [5, 58]	25.83	85.52	74.23