Two-Stage Adaptive Object Scene Flow Using Hybrid CNN-CRF Model

Congcong Li, Haoyu Ma, Qingmin Liao*

Department of Electronic Engineering
Tsinghua Shenzhen International Graduate School, Tsinghua University

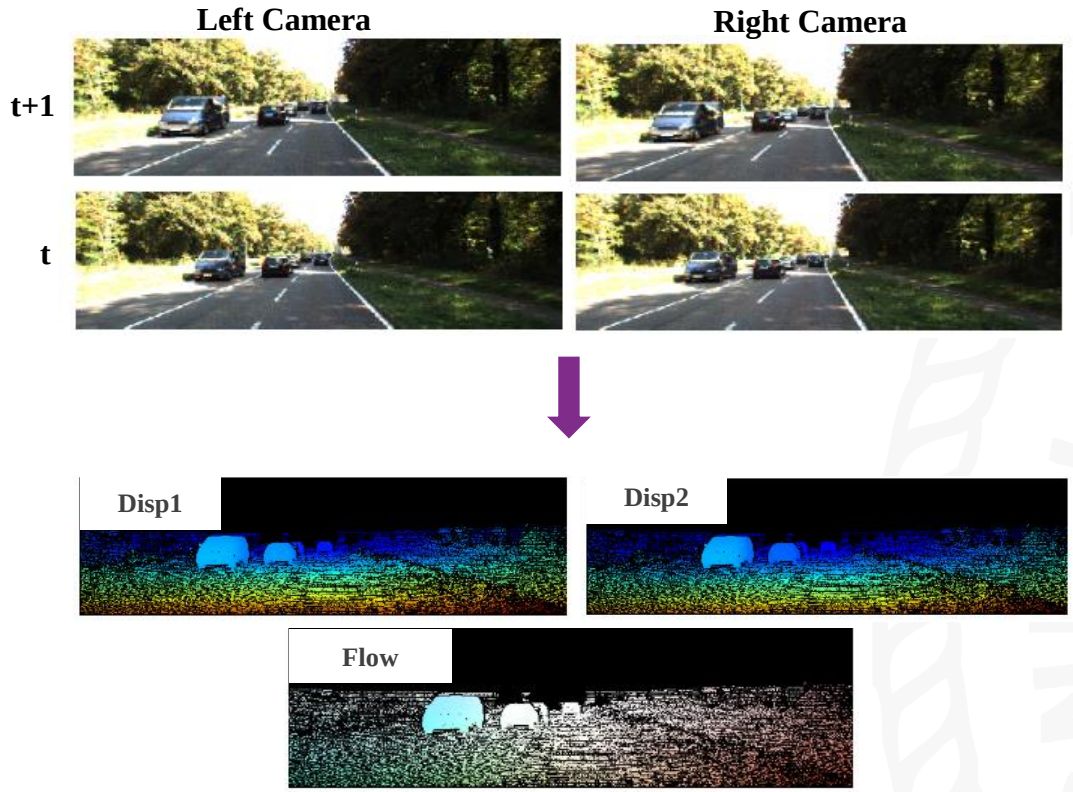
Outline

- Introduction
- Our Method
- Experimental Results
- Conclusion

Introduction

Scene Flow

"Scene flow is a dense three-dimensional vector field defined for each point on every surface in the scene." [1]



RGB image sequences (two temporally adjacent stereo image pairs of a calibrated camera)

[1] S. Vedula, P. Rander, R. Collins, and T. Kanade, "Three-dimensional scene flow," IEEE transactions on pattern analysis and machine intelligence, vol. 27, no. 3, pp. 475–480, 2005.

Introduction

Related Work

Segment Based

- sensitive to illumination change and large displacements
- computationally complex and heavy time-consuming

CNN Based

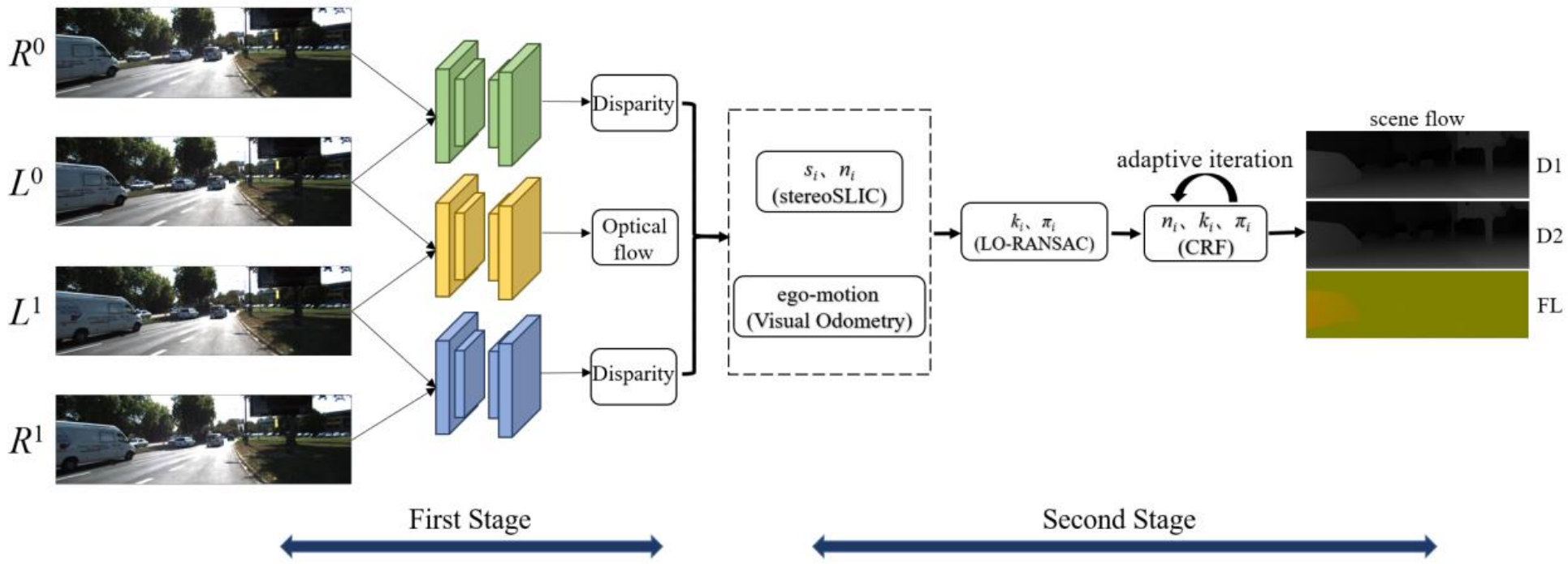
- not good enough to employ global context information to model interactions between predictions directly
- limited by the scarce datasets in the real scenarios

Ours

- ✓ propose a two-stage Adaptive Object Scene Flow estimation method using a hybrid CNN-CRF model (ACOSF), which combines the effectiveness of CNN and modelling ability of CRF
- ✓ employ adaptive iteration and high-quality pixel selection to balance the computational efficiency and accuracy

Our Method—ACOSF

Hybrid CNN-CRF Model (Two-Stage)



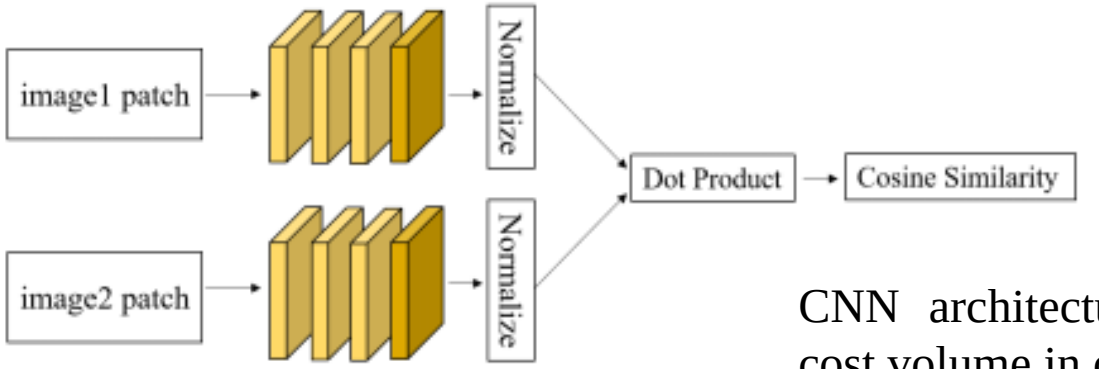
In the first stage, we use CNNs to obtain initial disparity and optical flow estimation. Then we integrate the initial results into a CRF-based model.

Our Method

Depth Information—Disparity

- We leverage the pyramid stereo matching network (PSM-Net) [1] to initialize disparity estimation, which exploits high-quality features to find correspondences.
- Stacked hourglass network in [1] is modified — Considering that the disparity maps in the first two hourglass networks are not utilized during inference, these auxiliary outputs can be removed to reduce the computational cost.

2D Optical Flow Estimation



CNN architecture for calculating the cost volume in optical flow estimation

[1] J.-R. Chang and Y.-S. Chen, "Pyramid stereo matching network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5410–5418.

Our Method

Tsinghua University

CRF Model for Scene Flow

- In the second stage, we over-segment the reference image L^0. And we follow the assumption[1] that there are a finite number of traffic participants moving rigidly and independently. OSF model[1] is adopted to infer the full scene flow.

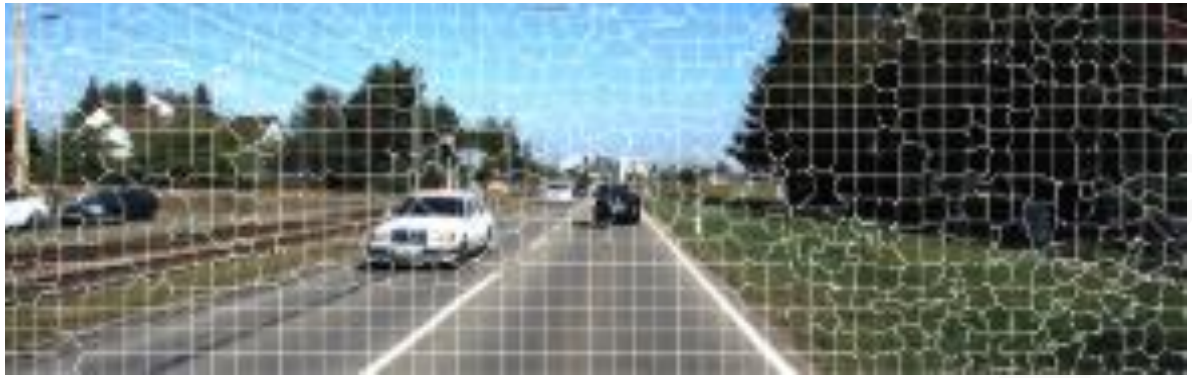
- Each planar region B_i in the image is allocated to superpixel $s_i \epsilon S$, which is described by a random variable $v_i = (k_i, s_i)^{\mathrm{T}}$. Each object O_k is associated with a variable $\pi_i \in SE(3)$ describing its rigid motion.

$$E(v,\pi) = \sum_{s_i \in S} \underbrace{\varphi_i(v_i,\pi)}_{\text{data}} + \sum_{s_i \sim s_j} \underbrace{\psi_{ij}(v_i,v_j)}_{\text{smoothness}}$$

[1] M. Menze and A. Geiger, “Object scene flow for autonomous vehicles,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3061 3070.

Our Method

Efficiency



- We make use of high-confidence matching obtained in the first stage. The algorithm samples a small number of pixels to construct the cost volume instead of all pixels in the region B_i.
- Our ACOSF can dynamically adjust the number of iterations n in the MP-PBP to suit different scenarios by comparing the continuous variation of the energy function with the pre-set threshold T, which makes the iteration number different for various scenes.

Experimental Results

Tsinghua University

Comparison with state-of-the-art

Method	D1			D2			Fl			SF			Run Time
	bg	fg	all	bg	fg	all	bg	fg	all	bg	fg	all	
DRISF [33]	**2.16**	**4.49**	**2.55**	**2.90**	**9.73**	**4.04**	**3.59**	**10.40**	**4.73**	**4.39**	**15.94**	**6.31**	0.75s(G)
ACOSF(ours)	**2.79**	7.56	3.58	**3.82**	12.74	**5.31**	**4.56**	12.00	**5.79**	**5.61**	19.38	**7.90**	5min(C)
ISF [12]	4.12	6.17	4.46	4.88	11.34	5.95	5.40	**10.29**	6.22	6.58	**15.63**	8.08	10min(C)
PRSM* [29]	3.02	10.52	4.27	5.13	15.11	6.79	5.33	13.40	6.68	6.61	20.79	8.97	5min(C)
OSF+TC* [31]	4.11	9.64	5.03	5.18	15.12	6.84	5.76	13.31	7.02	7.08	20.03	9.23	50min(C)
SSF [32]	3.55	8.75	4.42	4.94	17.48	7.02	5.63	14.71	7.14	7.18	24.58	10.07	5min(C)
OSF [11]	4.54	12.03	5.79	5.45	19.41	7.77	5.62	18.92	7.83	7.01	26.34	10.23	50min(C)
DWARF [18]	3.20	**3.94**	**3.33**	6.21	**9.38**	6.73	9.80	13.37	10.39	11.72	18.06	12.78	0.14-1.43s(G)
PWOC-3D [34]	4.19	9.82	5.13	7.21	14.73	8.46	12.40	15.78	12.96	14.30	22.66	15.69	0.13s(G)
CSF [43]	4.57	13.04	5.98	7.92	20.76	10.06	10.40	25.78	12.96	12.21	33.21	15.71	80s(C)

- ACOSF is demonstrated to compare favorably to the state-of-the-art and ranks 2nd in the leaderboard.
- The scene flow error of the proposed method is 23% lower and the running time is 10 times faster by comparison with OSF on the test set.

Experimental Results

Comparison with state-of-the-art

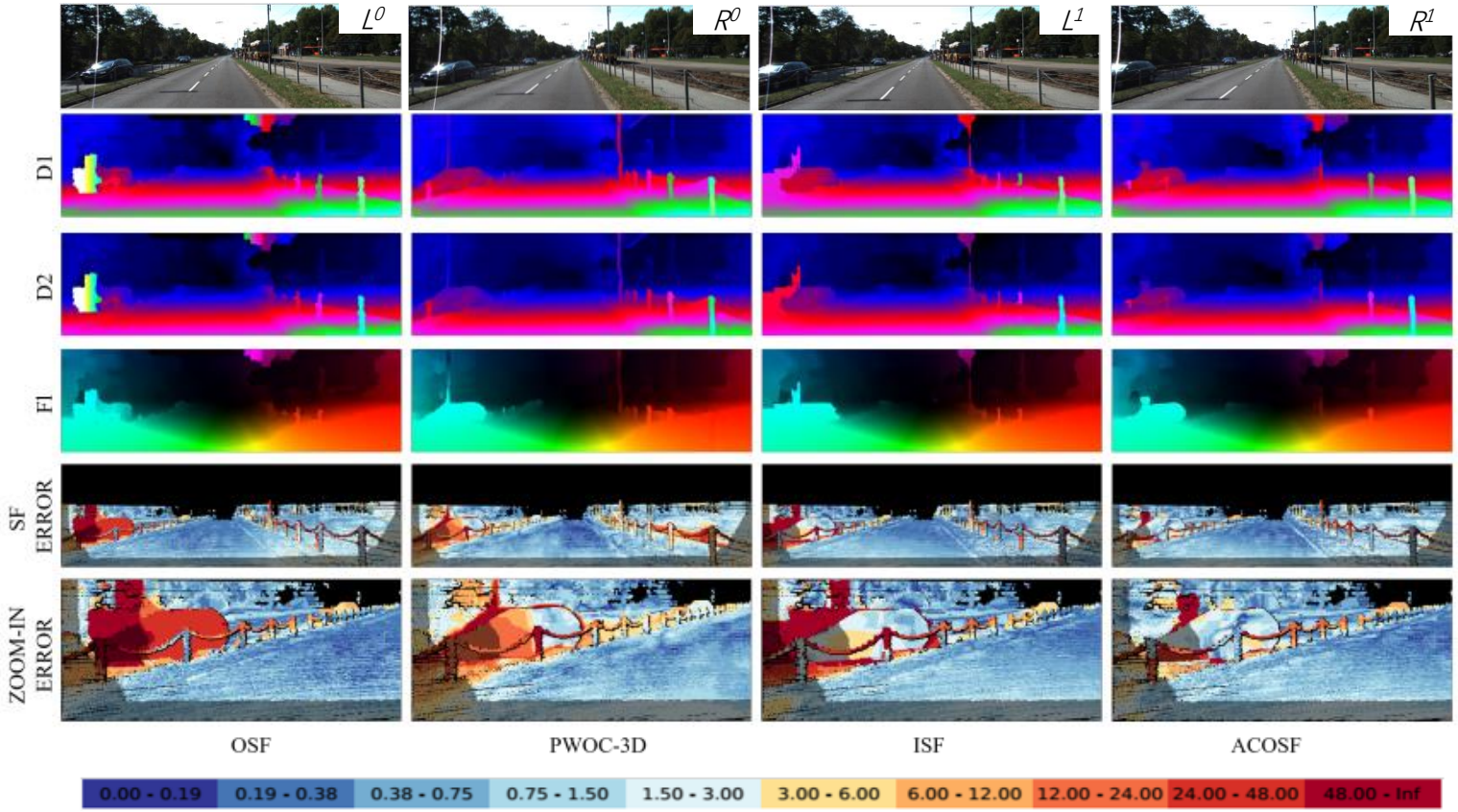


Fig. 4. **Comparison of visual results on the KITTI test set (best viewed in color).** The proposed method is more robust due to high-quality features obtained by CNN. Top-to-bottom in the upper and lower part respectively: reference image L^0, disparities in the first and second frame D1, D2, optical flow Fl, scene flow error and enlarged view of the car in the error map. The color in scene flow error map encodes the error from low (blue) to high (red) which is depicted in the bottom legend. Redder indicates higher error.

- We visualize a few scene flow results on the KITTI test set.
- ACOSF produces better disparity and optical flow map.
- The error map demonstrates that our method is more robust to complex brightness in the surroundings.

Conclusion

Tsinghua University

- We propose a two-stage Adaptive Object Scene Flow estimation method using a **hybrid CNN-CRF model**, which benefits from high-quality features and the structured modelling capability.

- With high-quality features extracted by CNNs, we obtain **robust initial solution** in the first stage, which is of benefit to optimize and infer the full scene flow. And we employ **high-quality pixel selection** to construct the cost volume in the second stage.

- We design a new optimization technique termed as **adaptive iteration** to make the algorithm flexible for various scenes.

- Our approach is demonstrated to compare favorably to the state-of-the-art and **ranks 2nd** in the challenging KITTI scene flow benchmark leaderboard.

Thanks for your attention!

Two-Stage Adaptive Object Scene Flow Using Hybrid CNN-CRF Model

Congcong Li, Haoyu Ma, Qingmin Liao

Email: licc18@mails.tsinghua.edu.cn