Unsupervised Deep Epipolar Flow for Stationary or Dynamic Scenes –Supplemental Material

Yiran Zhong[1,4,5], Pan Ji[2], Jianyuan Wang[1,3], Yuchao Dai[3], Hongdong Li[1,4]
[1]Australian National University, [2]NEC Labs America,
[3]Northwestern Polytechnical University, [4]ACRV, [5]Data61 CSIRO
{yiran.zhong, hongdong.li}@anu.edu.au, panji@nec-labs.com, daiyuchao@nwpu.edu.cn

Abstract

In this supplemental material, we provide more qualitative results of optical flow estimation on the KITTI VO dataset. Furthermore, on stationary scenes, we compare our method with the conventional RANSAC method [2] in fundamental matrix estimation on a subset of the KITTI 2012 training set.

1. Fundamental Matrix Recovery

Our method has the potential to recover a fundamental matrix for a stationary scene. To evaluate our recovered fundamental matrix, we select several stationary scenes from the KITTI 2012 training dataset and compare our results with the fundamental matrices computed from ground truth optical flows and the fundamental matrices computed from sparse feature correspondences. To compute fundamental matrix from ground truth optical flow, we first convert a ground truth flow into matching point pairs and then use the normalized eight-point algorithm to compute the corresponding fundamental matrix. For sparse feature correspondences based fundamental matrix, we detect SURF [1] features and perform feature matching. The fundamental matrix is computed with RANSAC + normalized eight-point algorithm. We use MATLAB's inbuilt function *estimateFundamentalMatrix*.

In Fig. 1, we illustrate our fundamental matrix estimation performance under these two conditions. We report our results based on the Gold Standard Method [3]. The left column visualizes the epipolar line estimation from a small forward movement while the right one from a large turn. Our baseline model, which only uses image warping constraint, shows competitive performance in small motions but has relatively large errors in large motions. Our methods that enforce the epipolar constraint, *i.e.*, Our-F, Our-low-rank and Our-sub, share similar performance, with Our-sub achieving the lowest re-projection error. Table 1 provides quantitative results, comparing our recovered $\mathbf{F}$ against the result of RANSAC [2].

References

[1] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006. 1

[2] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981. 1

[3] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 1

	Ground truth	RANSAC	Our-warp	Our-F	Our-low-rank	Our-sub
Re-projection error (px)	0.0104	0.3608	1.3557	1.0401	0.9753	0.9448

Table 1. **Average re-projection error over five testing pairs.** We evaluated various methods of computing fundamental matrix on the five randomly selected pairs from the KITTI 2012 training set. The error is computed over all the valid pixels in the ground truth optical flows.

Figure 1. **Epipolar lines of a random subset of inliers.** We randomly select 300 points from ground truth flow as a subset of inliers and visualize their epipolar lines. The ground truth $\mathbf{F}$ is computed from ground truth flow. The RANSAC represents the $\mathbf{F}$ that computed from normalized eight point algorithm with RANSAC (for 5000 iterations). We illustrate our $\mathbf{F}$ estimations of each method using Our-warp, Our-F, Our-low-rank and Our-sub. The camera in the left scene moves forward while the right one turns sharply.