Scene flow estimation based on stereo sequences is a comprehensive task relevant to disparity and optical flow. Some existing methods are time-consuming and often fail in the presence of reflective surfaces. In this paper, we propose a two-stage adaptive object scene flow estimation method using a hybrid CNN-CRF model (ACOSF), which benefits from high-quality features and the structured modelling capability. Meanwhile, in order to balance the computational efficiency and accuracy, we employ adaptive iteration for energy function optimization, which is flexible and efficient for various scenes. Besides, we utilize high-quality pixel selection to reduce the computation time with only a slight decrease in accuracy. Our method achieves competitive results with the state-of-the-art, which ranks second on the challenging KITTI 2015 scene flow benchmark.

Our ACOSF mainly consists of two stages based on the hybrid CNN-CRF model. 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.

3D Geometry and 2D Optical Flow Estimation

In the first stage, the convolutional neural networks are used to obtain the initial disparity and optical flow estimation, which are powerful to extract high-quality features for matching and searching correspondences.

CRF Model for Scene Flow

In the second stage, we over-segment the reference image $L^0$. And we follow the assumption in [1] that there are a finite number of traffic participants moving rigidly. Each planar region $B_i$ in the image is allocated to superpixel $s_i \in S$, which is described by a random variable $v_i = (n_i, k_i)^T$. Each object $O_k$ is associated with a variable $p_k \in SE(3)$ describing its rigid motion.

$$E(v, \pi) = \sum_{s_i \in S} \psi^v(v_i, \pi) + \sum_{\kappa \sim s_i} \psi^\pi(v_i, v_j)$$

Efficiency

1) To balance the accuracy and computational 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 $B_i$.
2) 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 $T$.

References