Multi-staged pipeline framework

We estimate disparity $D$, camera motion $P$, moving-object mask $S$, and moving-object flow $f_{non}$ (non-rigid flow) by implicitly minimizing image residual

$$ E(D, P, S, f_{non}) = \sum \| (p) - D_{\text{des}}(w(p; D, P, S, f_{non})) \| $$

using bimodal warping $w(p; D, P, S, f_{non}) = \begin{cases} p + f_{non}(p) & \text{if } \sigma(p) = \text{foreground} \\ p + f_{rig}(D, P) & \text{if } \sigma(p) = \text{background} \end{cases}$

- **Input**
  - Binocular stereo
  - Visual odometry
  - Epipolar stereo

- **Output**
  - Fully rigid warping $w(p; D, P)$
  - Partly non-rigid warping $w(p; D, P, S, f_{non})$

**Binocular stereo** uses SGM to get an initial disparity map.

**Visual odometry** estimates camera motion by minimizing

$$ \min \sum w_r \| D_{\text{des}}(w_r(p; D, P)) \| $$

We downweight moving object regions by $w_r$ predicted by previous $(S, f_{non})$.

**Epipolar stereo** refines disparity using temporarily adjacent frames. We blend left-right matching costs with matching costs for four adjacent frames.

**Initial segmentation** finds moving object regions. We use GrabCut with image residual as soft seeds for moving foreground.

**Image residual**

**Optical flow** estimates 2D flow map for only the predicted moving object regions. We use the SGM algorithm.

**Flow fusion** combines rigid and non-rigid flow proposals by a fusion move.

**Experiments**

KITTI 2015 stereo scene flow benchmark (in November 2016)

- **Improvements by epipolar stereo**
- **Evaluation on Sintel dataset**

**Comparison with state-of-the-art methods (PRSM, OSF) on Sintel dataset**

- Our method is better
- Our method is worse

**Fast Multi-frame Stereo Scene Flow with Motion Segmentation**

Tatsunori Tanai (RIKEN AIP/Univ. of Tokyo)  Sudipta N. Sinha (Microsoft Research)  Yoichi Sato (Univ. of Tokyo)

**Contributions**

Unified framework where multiple tasks benefit from each other

- **Optical flow**: 2D flow motion for rigid background (rigid flow) is recovered parametrically using known depth and camera motion, reducing computational burden of general (non-rigid) optical flow.

- **Stereo**: Given camera motion, disparity at left-right occluded regions is improved via multi-view stereo on consecutive frames.

- **Motion segmentation**: The segmentation mask is a byproduct of our flow estimation that fuses non-rigid and rigid flow maps.

- **Visual odometry**: Camera motion estimates are recovered more robustly by utilizing the moving object mask information.

In contrast to existing joint methods

- We decompose the task into several simple optimization problems, rather than directly optimizing a single complex function.