**Motivation:** Scene flow estimation often fails in the presence of large displacements or reflective surfaces, e.g., the front wheel in the first frame appears similar to the back wheel in the second frame.

**Goal:** To study the impact of three levels of recognition granularity on scene flow.

**Level of Recognition Granularity**

- Fine-grained geometric recognition:
  - Object coordinates: unique geometric labels of points on the object’s surface in its local coordinate space.
  - Superpixels inherit their rigid motion parameters from the associated object.
  - Bounding boxes & instance masks: provide powerful cues for object coordinates.

- Semantic grouping:
  - Detection and instance segmentation can provide powerful cues to identify potential dynamic objects.
  - Pools grouped together are likely to move as a single rigid object in the case of vehicles.

- Process:
  - Particle sampling at each iteration.
  - We use max-product particle BP to jointly infer shape and motion parameters with TRW-S for the inner loop.

**Representation**

- Input:
  - Parameters, 
  - Geometry variables: from a Gaussian distribution around the previous MAP estimate.

- Model:
  - Geometry variables: from a Gaussian distribution around the previous MAP estimate.
  - Instance term: encourages coherence of adjacent superpixels in terms of depth, orientation and motion. It decomposes as:
    \[
    \psi_I(q) = \sum_{j \in O} \sum_{p \in \mathcal{S}_j} \sum_{s \in \mathcal{S}} \psi_{ISF}(s) \chi_{v}(s) \psi_{osc}(s, q).
    \]

- Inference:
  - We use man-product particle BP to jointly infer shape and motion parameters with TRW-S for the inner loop.

**Experimental Results**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. PSNR</td>
<td>35.62</td>
<td>35.52</td>
<td>35.50</td>
</tr>
<tr>
<td>Avg. SSIM</td>
<td>0.85</td>
<td>0.84</td>
<td>0.83</td>
</tr>
</tbody>
</table>

**References**

