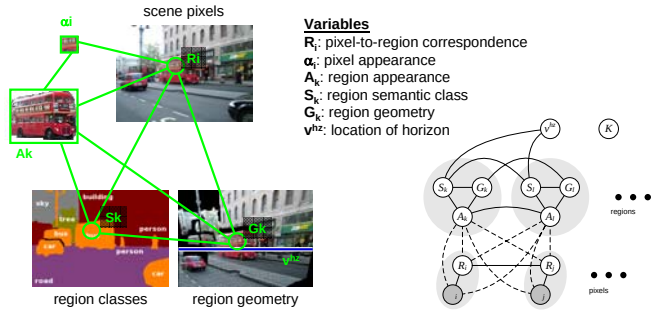
Decomposing a Scene into Geometric and Semantically Consistent Regions

Stephen Gould (Advisor: Daphne Koller)

1 Overview

- High-level scene understanding involves reasoning about objects, regions and the 3D relationships between them.
- This requires a decomposition of the scene from pixels into geometric and semantically consistent regions.
- We present a model for decomposing a scene based on a unified **energy function** which measures the "quality" of a decomposition.
- We show how this energy function can be learned and propose an efficient inference algorithm.
- By understanding regions, we can then start to address other vision tasks in a more satisfying way.

2 Scene Decomposition



3 Energy Function

$$E(\mathbf{R}, \mathbf{S}, \mathbf{G}, \mathbf{A}, v^{hz}, K \mid I, \theta) =$$

$\psi^{horizon}(S_k, G_k, v^{hz})$ + $\psi^{region}(S_k, A_k)$ + $\psi^{boundary}(R_i, R_j)$ + $\psi^{pair}(S_k, S_l)$

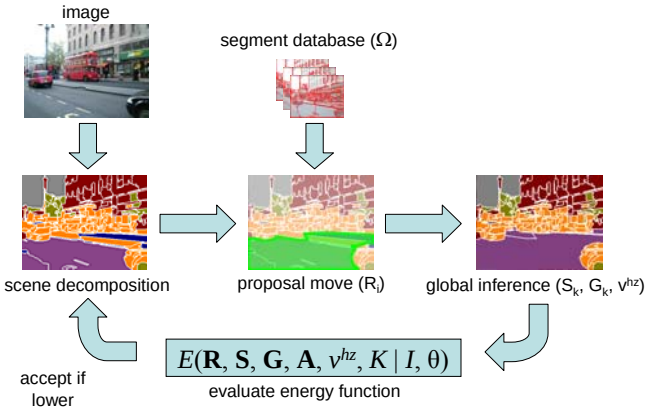
Horizon Term e.g. "sky" above horizon

Region Term e.g. consistent appearance

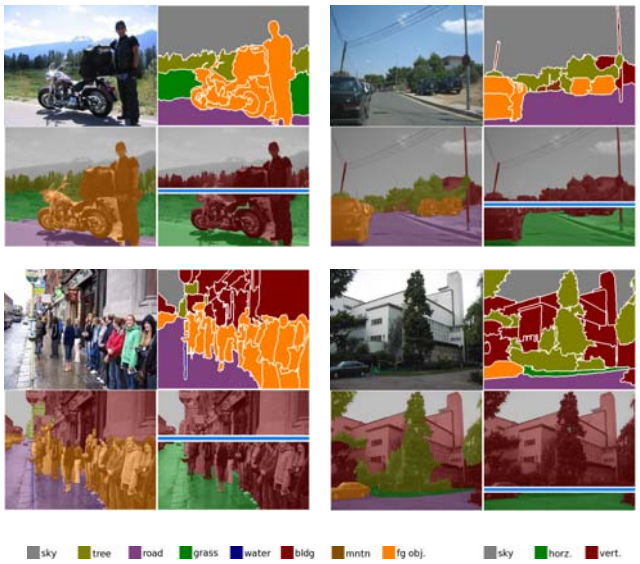
Boundary Term

Pairwise Term e.g. roads rarely next to water

4 MAP Inference



5 Example Results



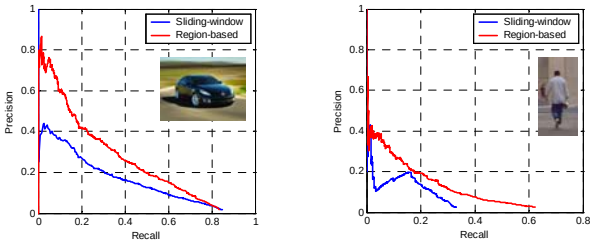
CLASS	Mean	Std
Pixelwise	74.3	0.80
Region-based	**76.4**	1.22

GEOMETRY	Mean	Std
Pixelwise	89.1	0.73
Region-based	**90.1**	0.56

6 Applications

- **3D Reconstruction**

- **Object Detection**



7 More Results