

## 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

scene pixels

**Variables**  
 $R_i$ : pixel-to-region correspondence  
 $\alpha_i$ : pixel appearance  
 $A_k$ : region appearance  
 $S_k$ : region semantic class  
 $G_k$ : region geometry  
 $v^{hiz}$ : location of horizon

region classes      region geometry

## 3 Energy Function

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

$\psi^{horizon}(S_k, G_k, v^{hiz})$   
  
**Horizon Term**  
 e.g. “sky” above horizon

$\psi^{region}(S_k, A_k)$   
  
**Region Term**  
 e.g. consistent appearance

$\psi^{boundary}(R_i, R_j)$   
  
**Boundary Term**

$\psi^{pair}(S_{k_1}, S_{k_2})$   
  
**Pairwise Term**  
 e.g. roads rarely next to water

## 4 MAP Inference

image

segment database ( $\Omega$ )

scene decomposition

proposal move ( $R_i$ )

global inference ( $S_k, G_k, v^{hiz}$ )

accept if lower

$E(\mathbf{R}, \mathbf{S}, \mathbf{G}, \mathbf{A}, v^{hiz}, K | I, \theta)$   
evaluate energy function

## 5 Example Results

■ sky    ■ tree    ■ road    ■ grass    ■ water    ■ bldg    ■ mntn    ■ fg obj.    ■ sky    ■ horiz.    ■ vert.

CLASS	Mean	Std
Pixelwise	74.3	0.80
Region-based	<b>76.4</b>	1.22

GEOMETRY	Mean	Std
Pixelwise	89.1	0.73
Region-based	<b>90.1</b>	0.56

## 6 Applications

- 3D Reconstruction**

- Object Detection**

## 7 More Results