Putting Objects in Perspective

Derek Hoiem Alexei A. Efros Martial Herbert

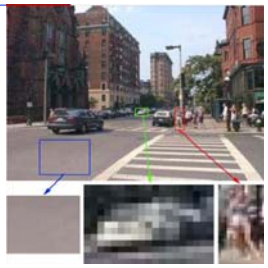
Presented by:

Inwoo Ha

Yong Jae Lee

Motivation

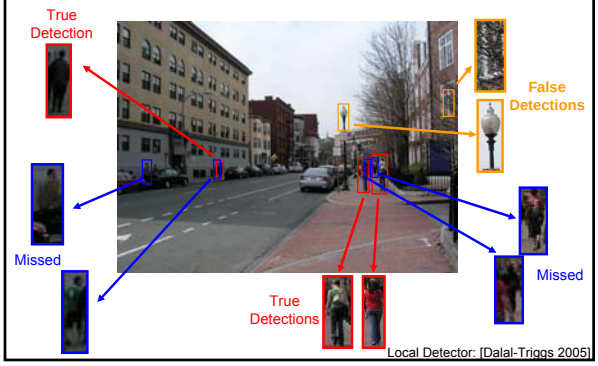
- How to recognize the car-like blob?
- General object recognition cannot be solved locally.
- The interpretation of the entire image is required.
- Real relationships are 3D.
  - It's sitting on the road.
  - It's the "right"size, relative to other objects in the scene (cars, buildings, and pedestrians, etc.)



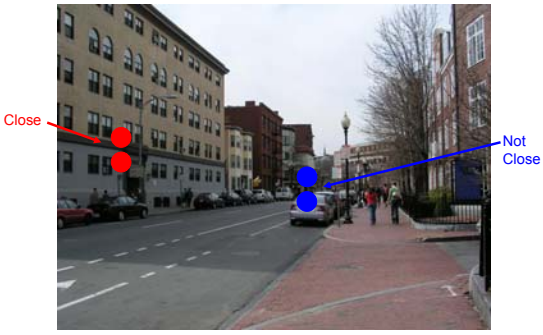
Understanding an Image

Today: Local and Independent

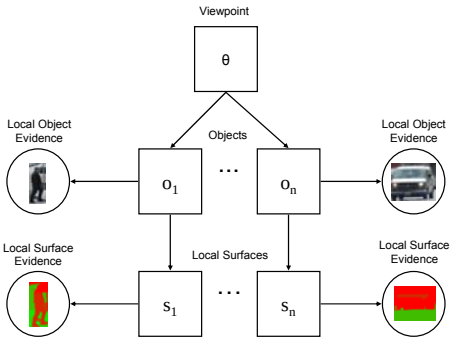
Local Object Detection



Real Relationships are 3D



Inference over Tree Easy with BP



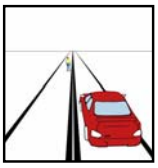
Viewpoint

- Assume the objects all lie on the ground plane.
- Then, we only need two variables:
  - Horizon position
  - Camera height
- How do we determine these parameters from a single image?
- Start with simple priori models.
  - Horizon position: 0.50
  - Camera height: 1.67

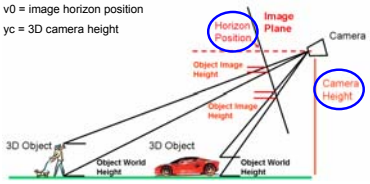
Object Size in the Image

$$y_i = \frac{h_i y_c}{v_i - v_0}$$

yi = 3D object height
hi = image object height
vi = image object bottom position
v0 = image horizon position
yc = 3D camera height



Image



World

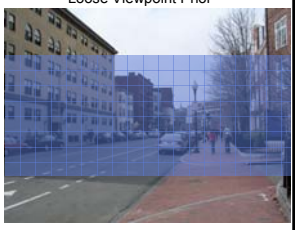
Objects

- Use window-based object detector at each position and scale.
- Can contain several dozen object candidates.
- An object's height depends on its position when given the viewpoint.
- The object candidates are constrained with the object's height.

Object Size ↔ Camera Viewpoint

Input Image

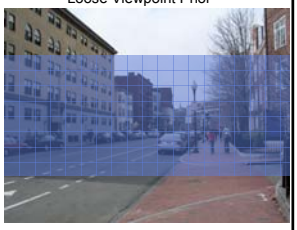
Loose Viewpoint Prior



Object Size ↔ Camera Viewpoint

Input Image

Loose Viewpoint Prior



Object Size ↔ Camera Viewpoint

Object Position/Sizes ↔ Viewpoint

Object Size ↔ Camera Viewpoint

Object Position/Sizes → Viewpoint

Object Size ↔ Camera Viewpoint

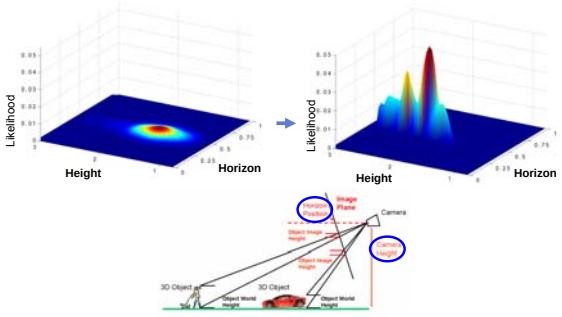
Object Position/Sizes ↔ Viewpoint

Object Size ↔ Camera Viewpoint

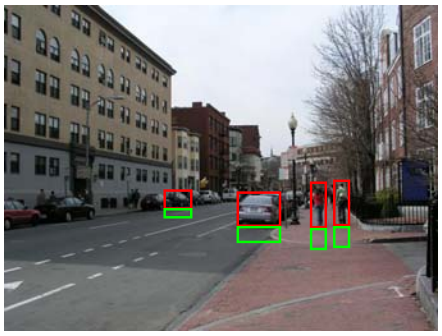
Object Position/Sizes ↔ Viewpoint

Viewpoint estimation

Viewpoint Prior → Viewpoint Final

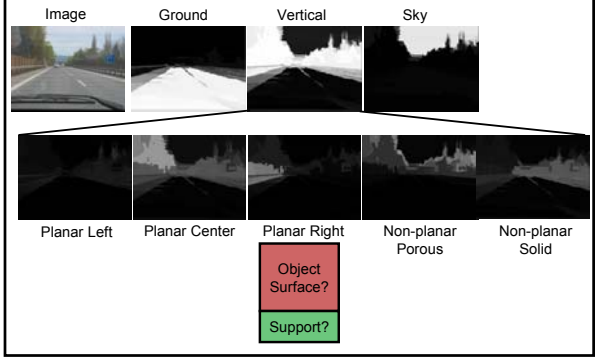


Object Support



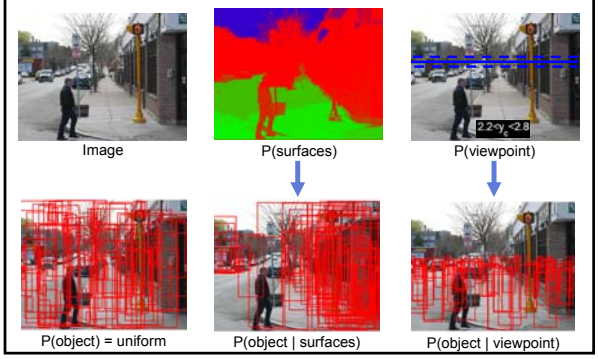
Surface Estimation

[Hoiem, Efros, Hebert ICCV 2005]

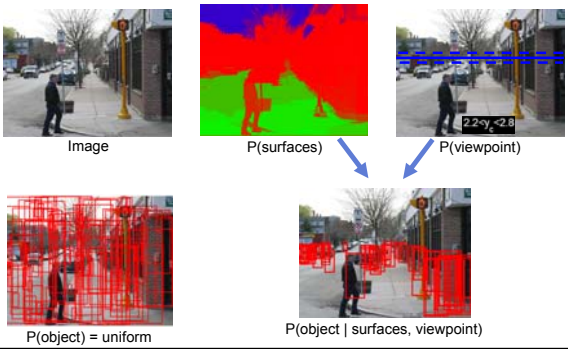


Surface Geometry

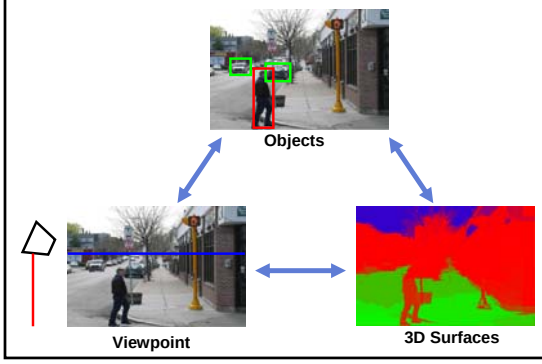
What does surface and viewpoint say about objects?



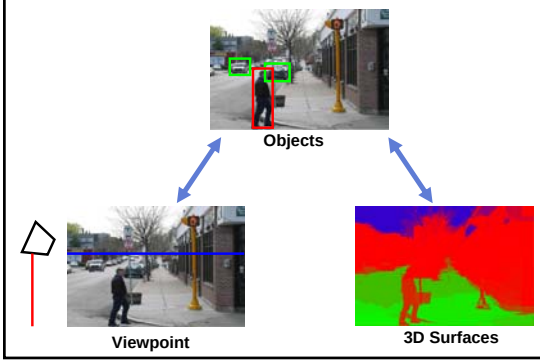
What does surface and viewpoint say about objects?



Scene Parts Are All Interconnected



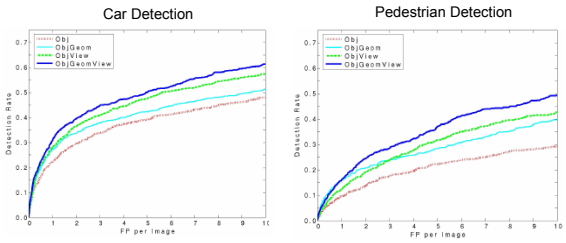
Approximate Model



Experiments

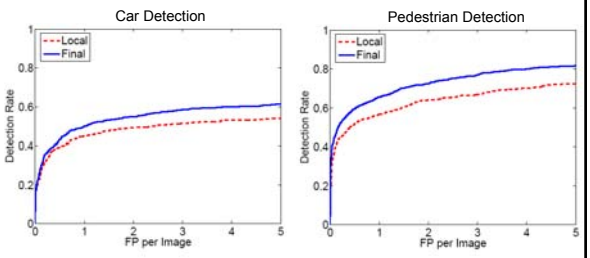
- Testing with LabelMe dataset: 422 images
  - 923 Cars at least 14 pixels tall
  - 720 Peds at least 36 pixels tall

Each piece of evidence improves performance



Local Detector from [Murphy-Torralba-Freeman 2003]

Can be used with any detector



Local Detector: [Dalal-Triggs 2005] (SVM-based)

Quantitative Results

Improved Horizon Estimation

	Mean	Median
Prior	10.0%	8.5%
+Obj	7.5%	4.5%
+ObjGeom	7.0%	3.8%

Mean/median absolute error (as % of image height) for horizon estimates

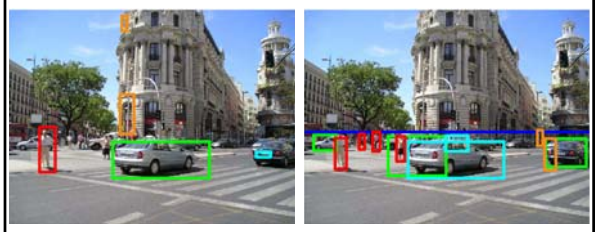
More is Better

	Horizon	Cars (FP)		Ped (FP)	
Car	7.3%	5.6	7.4	—	—
Ped	5.0%	—	—	12.4	13.7
Car+Ped	3.8%	5.0	6.6	11.0	10.7

Median absolute error of horizon, # of FP per image at 50% detection rate

Qualitative Results

Car: TP / FP Ped: TP / FP



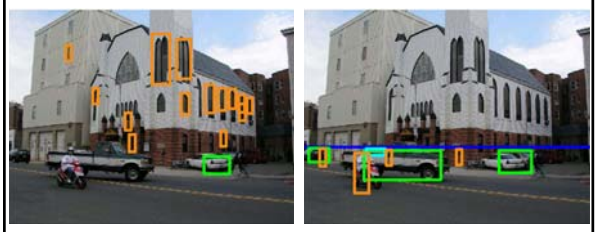
Initial: 2 TP / 3 FP

Final: 7 TP / 4 FP

Local Detector from [Murphy-Torralba-Freeman 2003]

Qualitative Results

Car: TP / FP Ped: TP / FP



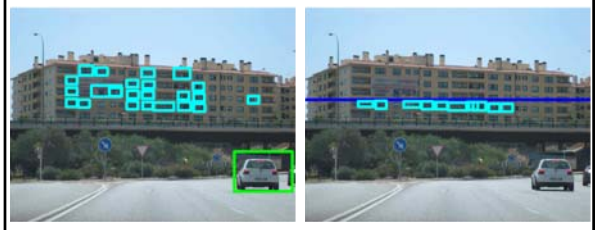
Initial: 1 TP / 14 FP

Final: 3 TP / 5 FP

Local Detector from [Murphy-Torralba-Freeman 2003]

Qualitative Results

Car: TP / FP Ped: TP / FP



Initial: 1 TP / 23 FP

Final: 0 TP / 10 FP

Local Detector from [Murphy-Torralba-Freeman 2003]

Summary

- A statistical framework that allows *simultaneous* inference of objects, surface geometry, and camera viewpoint

Conclusion

- Image understanding is a 3D problem
  - Recognition, Scene Geometry, Segmentation
  - Must be solved jointly
  - In 3D world in which we live, not in 2D world of the image

References

- D. Hoiem, A. Efros, and M. Hebert, “Putting Objects in Perspective”, *CVPR 2006*.
- D. Hoiem, A.A. Efros, and M. Hebert, "Geometric Context from a Single Image", *ICCV* 2005.
- Presentation Slides: “Putting Objects in Perspective” Prepared for *CVPR 2006*.