Recovering Occlusion Boundaries from a Single Image

Zhang, Liqiang

June 16, 2018

Abstract

Occlusion reasoning, necessary for tasks such as navigation and object search, is an important aspect of everyday life and a fundamental problem in computer vision. The amazing ability of humans to reason about occlusions from one image is based on an intrinsically 3D interpretation. In this paper, the goal is to recover the occlusion boundaries and depth ordering of free-standing structures in the scene. This approach is to learn to identify and label occlusion boundaries using the traditional edge and region cues together with 3D surface and depth cues. Since some of these cues require good spatial support, the author gradually create larger regions and use them to improve inference over the boundaries.

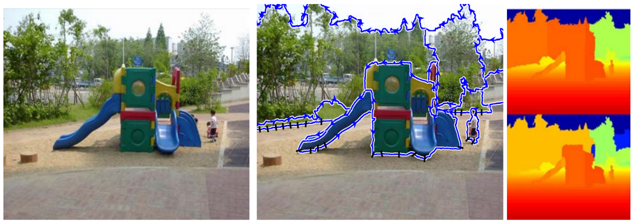

Figure 1. Given an image (left), we recover occlusion boundaries (center) and infer a range of possible depths (right) that are consistent with the occlusion relationships. In the center, blue lines denote occlusion boundary estimates, arrows indicate which region (left) is in front, and black hatch marks show where an object is thought to contact the ground. On the right, we display the minimum and maximum depth estimates (red = close, blue = far).

1. Introduction

What makes scene understanding different from other image processing tasks, such as medical or aerial image analysis, is the notion that the image is not a direct representation, but merely a projection of the 3D scene. One major consequence of this projection is occlusion C the concept that two objects that are spatially separated in the 3D world might interfere with each other in the projected 2D image plane. Consider the scene in Fig. 1: nearly every object is partially occluded by some other object, and each occludes part of the ground. Yet, despite their pervasiveness, occlusions have too often been ignored [2]. In computer vision, the study of occlusion reasoning has been largely confined to the context of stereo, motion and other multiview problems.

In this paper, the author argue that occlusion reasoning lies at the core of scene understanding and must be addressed explicitly [3]. Their goal is to recover the boundaries and depth ordering of prominent objects in sufficient detail to provide an accurate sense of depth. The greatest challenge is that objects are typically defined, not by homogeneity in appearance, but by physical connectedness. For example, in Fig. 1 the most prominent objects are the jungle gym, the boy, and the vegetation.

This approach is to learn models of occlusion based on both 2D perceptual cues and 3D surface and depth cues from a training set. We can then use those learned models to gradually infer the occlusion relationships, reasoning together about boundaries in the image and surfaces in the scene.

2. CRF Model for Occlusion Reasoning

This paper's CRF model allows joint inference over both boundary and surface labels, modeling boundary strength and continuity and enforcing closure and surface/boundary consistency. The author represent the model with a factor graph, in which the probability of the boundaries and surfaces is given by

$$P(labels|data) = \frac{1}{Z}\prod_{j}^{N_j}\phi_j\prod_{e}^{N_e}\gamma_e \tag{1}$$

where in shorthand notation we denote junction factor ϕ_j and surface factor γ_e, with N_j junctions and N_e boundaries in the graph and partition function Z.

The junction factors encode the likelihood of the label of each boundary according to the data, conditioned on its preceding boundary if there is one [4]. They also enforce closure and continuity. Though there are 27 possible labelings of boundary triplets, there are only five valid types of threejunctions, up to a permutation.

3. Experiments

	Edge/Region Cues	+ 3D Cues	with CRF
Iter 1	58.7%	71.7%	-
Iter 2	65.4%	75.6%	77.3%
Final	68.2%	77.1%	**79.9%**

Table 1. Figure/ground labeling accuracy results for using edge/region cues only, all cues (including 3D cues), and after performing inference using our CRF model (only unary likelihoods were used in the first iteration).

	Conservation	$\log_2$ Efficiency
Our Algorithm	**83.7%**	**-0.8**
Surface-Based	82.4%	-1.4
Ncuts	81.7%	-1.2

Table 2. The author outperform segmentations using only surface labels and an image-based normalized cuts algorithm [1] by using both surface and image cues together with boundary reasoning.

In Table 1, the author report the figure/ground classification accuracy. Accuracy is computed over all true occlusion boundaries, including those which are incorrectly classified as non-boundaries in testing. In Table 2, the author show that their algorithm achieves greater mean conservation and efficiency than both of the others. Also, only their algorithm provides figure/ground labels.

References

[1] T. L. Berg, A. C. Berg, and J. Shih. Automatic attribute discovery and characterization from noisy web data. In *ECCV*, 2010. 2

[2] M. J. Black and D. J. Fleet. Probabilistic detection and tracking of motion discontinuities. *IEEE TPAMI*, 38(3):231–245, 2000. 1

[3] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multiscale graph decomposition. In *CVPR*, 2005. 1

[4] E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky. Depth from familiar objects: A hierarchical model for 3D scenes. In *CVPR*, 2006. 1