Computer Vision and Image Understanding 117 (2013) 42–55

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Occlusion cues for image scene layering ☆

Xiaowu Chen, Qing Li *, Dongyue Zhao, Qinping Zhao

State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China

ARTICLE INFO

Article history:
Received 4 January 2011
Accepted 2 October 2012
Available online 23 October 2012

Keywords:
Human perception
Occlusion cues
Occlusion prediction
Layering

ABSTRACT

To bring computer vision closer to human vision, we attempt to enable computer to understand the occlusion relationship in an image. In this paper, we propose five low dimensional region-based occlusion cues inspired by the human perception of occlusion. These cues are semantic cue, position cue, compactness cue, shared boundary cue and junction cue. We apply these cues to predict the region-wise occlusion relationship in an image and infer the layer sequence of the image scene. A preference function, trained with samples consisting of these cues, is defined to predict the occlusion relationship in an image. Then we put all the occlusion predictions into the layering algorithm to infer the layer sequence of the image scene.

The experiments on rural, artificial and outdoor scene datasets show the effectiveness of our method for occlusion relationship prediction and image scene layering.

1. Introduction

It is well known that an image is not a direct representation, but a projection of the 3D world. The understanding of an image scene includes not only the comprehension of the regions presented in the 2D image plane, but also the 3D spatial layout of the regions in the real world. We humans can immediately grasp the spatial relationship of the scene. Our perception includes the immediately visible portions as well as the estimation of the entire space. It can be seen from Fig. 1 (the input image is taken from the LHI dataset [1]) that our perception includes not only the visible textures of the rhinoceros, grass, ground, tree and sky, but also includes the estimation of the spatial relationship. For example, the rhinoceros standing on the ground is in front of the tree and grass. The ultimate goal of computer vision is to provide the computer with the same spatial understanding so that it can see the world as humans do.

Modern computer vision techniques, such as image segmentation, object detection and recognition and depth estimation, can be used to recover a lot of useful information about the image. Although a great progress has been made, it still remains extremely challenging for current computer vision systems to understand scenes as humans do. When the real world is projected into the image plane, occlusions will frequently occur between the objects that are spatially separated. In fact, almost every object in the image is occluded by, and (or) occludes other objects. Understanding occlusion helps us to comprehend the 3D spatial relationship. Thus, to bring computer vision closer to human vision, we need to make the computer understand the occlusion relationship. Recently there are some works paying attention to this topic such as the 2.1 D sketch [2] and the occlusion boundary detection [3,4]. They mainly focus on the occlusion boundary identification, but ignore the recognition of the global image. In addition, there is still a large gap between human perception and occlusion reasoning. How to bridge this gap is our main concern in this paper.

Previous works on this topic used a wide variety of perception features to recover the occlusion relationship. Inspired by these works, we use occlusion features to express the human perception rules. There are many low-level and high-level features used in the occlusion reasoning, figure/ground assignment and depth ordering, such as texture features, color features and gestalt cues. In psychophysics and cognition, gestalt cues, including size, convexity, symmetry, parallelism, surroundedness and lower-region, are considered to be useful and important. Inspired by human perception, we propose five low dimensional region-based occlusion cues (referred to as ‘five cues’ below), which are semantic cue, position cue, compactness cue, shared boundary cue and junction cue.

In this paper, we describe our five cues and show their capabilities of occlusion relationship prediction. The experimental results show that our cues are efficient, and thus demonstrate that our cues can recover occlusion relationship to some extent. We apply all the occlusion predictions in an image to infer the layer sequence of the image scene. Our layer sequence indicates the layer partition of regions, and each region is assumed to be assigned a unique layer. Fig. 1 shows the overview of our method. Since our cues are based on the regions in the image, we first preprocess an input image to get its semantic label map. Our cues can work well regardless of how to get the semantic label map. Thus semantic

☆ This paper has been recommended for acceptance by J.K. Tsotsos.

* Corresponding author.
E-mail address: liqing@vrlab.buaa.edu.cn (Q. Li).

1077-3142/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cviu.2012.10.001

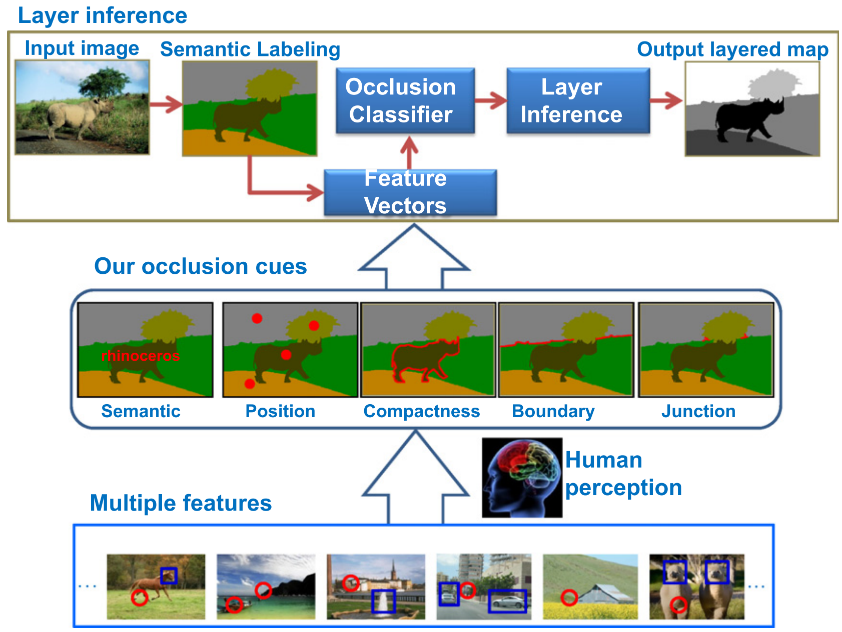

Fig. 1. The overview of our method. Inspired by the human occlusion perception, we propose five occlusion cues. We apply these cues to predict the occlusion relationship in an image and use the predictions to infer the layer sequence of the image scene. In the layer inference process, we first get the semantic label map of a given input image since our cues are based on the regions. Then we compute values of the occlusion cues to form feature vectors. Next, the trained occlusion classifier is used to predict the occlusion relationship. Finally, we put all the predictions of the input image into the layer inference algorithm and get the layer sequence which is visualized as a layer map with the blacker region in the front.

labeling work is not the main concern in this paper. In the section of experiments, we perform scene layering based on both the semantic label maps obtained manually and by semantic labeling methods, and both give promising results.

The main contributions of this paper include: (1) We propose five region-based cues to describe the occlusion relationship in an image, which are to some extent efficient to indicate the occlusion relationship in image scenes. (2) Then we analyze the interactions of these five cues as their importance varies in different datasets. (3) Finally, we give a selection scheme of these cues when they are applied to occlusion perception problem.

The remainder of this paper is organized as follows. We review the relevant literatures in Section 2. Our occlusion cues are described in Section 3. Section 4 introduces a previous semantic labeling work done by us and describes our layering inference procedure. To test the effectiveness of our cues and validate our results, three types of experiments are performed in Section 5. Experimental verification of our cues is demonstrated in Section 5.2, and followed by the analysis of the interactions between cues in Section 5.3. Scene layering results on multiple datasets are shown in Section 5.4. Section 6 is a brief conclusion.

2. Related works

The last decade has seen an increase of interest in occlusion reasoning. In computer vision, the study of occlusion reasoning has been largely confined to the context of stereo, motion and other multi-view problems [5–7]. For single-view tasks, the study of occlusion has been focused on how to recover the 3D information hidden in the 2D image plane, such as the 2.1D sketch [2], occlusion recovery and contour completion [8,9,3,10,11], image segmentation and depth recovery [4,12–16], and 3D layout recovery and modeling [17,18]. Nitzberg and Mumford [2] formulated the 2.1D sketch problem as an energy minimization problem. The goal of the 2.1D sketch is to recover the partial occluding order relation and to complete the occluded contour simultaneously. Moreover, it keeps multiple reasonable solutions accounting for the intrinsic ambiguity caused by occlusion. Wang et al. [5] proposed the concept of layered representation for image coding and motion analysis. Since then, there have been a certain number of research works on the 2.1D sketch problem.

Ren et al. [4] presented an approach for figure/ground assignment. Yang et al. [12] used the layer order to label the pixels in the image. Yu et al. [13] integrated occlusion cues with figure/ground segregation by using hierarchical Markov random field. Following the energy function defined by Nitzberg and Mumford [2], Esedoglu and March [15] proposed to segment an image with depth information but without detecting junctions. Hoiem et al. [3] proposed a method to recover occlusion boundaries by learning a CRF model. Liu et al. [16] performed a semantic segmentation of the scene and used semantic labels to guide the 3D reconstruction. Hedau et al. [17] used a parametric 3D 'box' to model the global room space, and introduced a structured learning algorithm to choose the set of parameters. Saxena et al. [18] created 3D models by using a MRF to infer a set of 'plane parameters' that can capture both the 3-D location and 3-D orientation of small homogeneous patches in an image.

In all these cases above, they either focus on the perceptual completion caused by occlusion, or on the application of occlusion reasoning such as segmentation and 3D modeling, but ignore the potential perceptual rules. We attempt to find some rules that can make computer reason occlusion relationship as humans do. Some of the above literatures have used the cues such as color and texture [3,16], surface layout [3], boundary [7,10], contour [8] and junction [19] to reason the occlusion or estimate the pixel depth in an image. Other literatures have focused on how to bridge the gap between computer vision and human perception [20–22]. They concentrate on multiple features or cues, such as shape, curvature and orientation, to obtain the region perception. According to these works, the occlusion cues may be a starting point to the addressing of our occlusion perception problem.