Int J Comput Vis (2012) 99:123–146
DOI 10.1007/s11263-012-0521-z

Shape-Based Object Detection via Boundary Structure Segmentation

Alexander Toshev · Ben Taskar · Kostas Daniilidis

Received: 3 July 2011 / Accepted: 23 February 2012 / Published online: 22 March 2012

Abstract We address the problem of object detection and segmentation using global holistic properties of object shape. Global shape representations are highly susceptible to clutter inevitably present in realistic images, and thus can be applied robustly only using a precise segmentation of the object. To this end, we propose a figure/ground segmentation method for extraction of image regions that resemble the global properties of a model boundary structure and are perceptually salient. Our shape representation, called the chordiogram, is based on geometric relationships of object boundary edges, while the perceptual saliency cues we use favor coherent regions distinct from the background. We formulate the segmentation problem as an integer quadratic program and use a semidefinite programming relaxation to solve it. The obtained solutions provide a segmentation of the object as well as a detection score used for object recognition. Our single-step approach achieves state-of-the-art performance on several object detection and segmentation benchmarks.

Keywords Shape representation · Shape matching · Object recognition and detection · Object segmentation

A. Toshev (✉)
Google Research, 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
e-mail: toshev@google.com

B. Taskar · K. Daniilidis
GRASP Lab, University of Pennsylvania, 3330 Walnut St, Philadelphia, PA 19104, USA

B. Taskar
e-mail: taskar@cis.upenn.edu

K. Daniilidis
e-mail: kostas@cis.upenn.edu

1 Introduction

A multitude of different object representations have been explored, ranging from texture and local features to region descriptors and object shape. Although local features based on image gradients and texture perform relatively well for some object classes, many classes are not modeled sufficiently by local descriptors. For objects primarily characterized by distinctive shape, local texture features typically provide weak descriptions. In this paper we focus on the problem of exploiting global shape properties for object detection. Moreover, we tightly couple these properties to object segmentation, which makes shape-based detection possible in cluttered scenes.

Shape is commonly defined in terms of the set of contours that describe the boundary of an object. In contrast to gradient- and texture-based representations, shape is more descriptive at a larger scale, ideally capturing the object of interest as a whole. This has been recognized by the Gestalt school of perception, which has established the principle of *holism* in visual perception (Palmer 1999; Koffka 1935). This principle suggests that an object should be perceived in its totality and not merely as an additive collection of individual parts. The essential goal of a holistic representation for object recognition is to capture not just the presence of object parts but also non-local relationships between these parts. In this work, our response to the mantra 'the whole is greater than the sum of its parts' is 'the whole is the sum of all the relationships between its parts', as we make precise below.

Some of the most notable holistic representations are based on global transforms, such as Fourier transform (Zhang and Lu 2003) or the Medial Axis Transform (Blum 1973). Unfortunately, such transforms assume a pre-segmented object shape as input. As a result, the above rep-