Int J Comput Vis (2013) 103:213–225
DOI 10.1007/s11263-012-0583-y

Inference Methods for CRFs with Co-occurrence Statistics

Ľubor Ladický · Chris Russell · Pushmeet Kohli · Philip H. S. Torr

Received: 21 January 2011 / Accepted: 30 September 2012 / Published online: 9 November 2012

Abstract The Markov and Conditional random fields (CRFs) used in computer vision typically model only local interactions between variables, as this is generally thought to be the only case that is computationally tractable. In this paper we consider a class of global potentials defined over all variables in the CRF. We show how they can be readily optimised using standard graph cut algorithms at little extra expense compared to a standard pairwise field. This result can be directly used for the problem of *class based image segmentation* which has seen increasing recent interest within computer vision. Here the aim is to assign a label to each pixel of a given image from a set of possible object classes. Typically these methods use random fields to model local interactions between pixels or super-pixels. One of the cues that helps recognition is global *object co-occurrence statistics*, a measure of which classes (such as chair or motorbike) are likely to occur in the same image together. There have been several approaches proposed to exploit this property, but all of them suffer from different limitations and typically carry a high computational cost, preventing their application on large images. We find that the new model we propose produces a significant improvement in the labelling compared to just using a pairwise model and that this improvement increases as the number of labels increases.

Keywords Conditional random fields · Object class segmentation · Optimization

Ľubor Ladický, Chris Russell contributed equally and have joint first authorship.

Ľ. Ladický (✉)
University of Oxford, Oxford, UK
e-mail: lubor@robots.ox.ac.uk

C. Russell
Queen Mary College, University of London, London, UK
e-mail: chrisr@eecs.qmul.ac.uk

P. Kohli
Microsoft Research, Cambridge, UK
e-mail: pkohli@microsoft.com

P. H. S. Torr
Oxford Brookes University, Oxford, UK
e-mail: philiptorr@brookes.ac.uk

1 Introduction

Class based image segmentation is a highly active area of computer vision research as shown by a spate of recent publications (Heitz and Koller 2008; Rabinovich et al. 2007; Shotton et al. 2006; Torralba et al. 2003; Yang et al. 2007). In this problem, every pixel of the image is assigned a choice of object class label, such as grass, person, or dining table. Formulating this problem probabilistically, in order to perform inference, is a difficult problem, as the cost or energy associated with any labelling of the image should take into account a variety of cues at different scales. A good labelling should take account of: low-level cues such as colour or texture (Shotton et al. 2006), that govern the labelling of single pixels; mid-level cues such as region continuity, symmetry (Ren et al. 2005) or shape (Borenstein and Malik 2006) that govern the assignment of regions within the image; and high-level statistics that encode inter-object relationships, such as which objects can occur together in a scene. This combination of cues makes for a multi-scale cost function that is difficult to optimise.

Current state of the art low-level approaches typically follow the methodology proposed in *Texton-boost* (Shotton et al. 2006), in which weakly predictive features such as colour, location, and texton response are used to learn a classifier which provides costs for a single pixel taking a particular