Int J Comput Vis (2012) 100:38–58
DOI 10.1007/s11263-012-0531-x

Minimizing Energies with Hierarchical Costs

Andrew Delong · Lena Gorelick · Olga Veksler · Yuri Boykov

Received: 15 October 2011 / Accepted: 21 April 2012 / Published online: 9 May 2012

Abstract Computer vision is full of problems elegantly expressed in terms of energy minimization. We characterize a class of energies with *hierarchical costs* and propose a novel *hierarchical fusion* algorithm. Hierarchical costs are natural for modeling an array of difficult problems. For example, in semantic segmentation one could rule out unlikely object combinations via hierarchical context. In geometric model estimation, one could penalize the number of unique model *families* in a solution, not just the number of models—a kind of hierarchical MDL criterion. Hierarchical fusion uses the well-known α-expansion algorithm as a subroutine, and offers a much better approximation bound in important cases.

Keywords Energy minimization · Hierarchical models · Graph cuts · Markov random fields (MRFs) · Segmentation

1 Introduction

Energy minimization is of strong practical and theoretical importance to computer vision. An energy expresses our criteria for a good solution—low energies are good, high energies are bad—independent of any algorithm. Algorithms are however hugely important in practice. Even for low-level vision problems we are confronted by energies that are computationally hard (often NP-hard) to minimize. As a consequence, a significant portion of computer vision research is dedicated to identifying energies that are useful and yet reasonably tractable. Our work is of precisely this nature.

Computer vision is full of 'labeling' problems cast as energy minimization. For example, the data to be labeled could be pixels, interest points, point correspondences, or mesh data such as from a range scanner. Depending on the application, the labels could be either semantic (object classes, types of tissue) or describe geometry/appearance (depth, orientation, shape, texture).

There are many labeling problems for which the labels naturally form groups. In computer vision, a recent trend is the use of 'context' to resolve ambiguities in object recognition (*e.g.* Choi et al. 2010; Ladický et al. 2010a; Zhou et al. 2011). The idea is that certain groups of labels are self-consistent because they tend to appear together, *e.g.* the {*car*, *road*, *sky*} labels all belong to the "outdoors" context, while {*table*, *chair*, *wall*} all belong to the "indoors" context. In computer graphics, one may wish to automatically classify the faces of a 3D mesh into semantic parts for the benefit of artists and animators (Kalogerakis et al. 2010). The part labels *arm*, *tail*, and *wheel* naturally belong to different groups based on their context (humanoid, quadruped, vehicle). In operations research, *facility location* can be cast as a labeling problem, and hierarchical variants have been studied (Svitkina and Tardos 2006; Sahin and Süral 2007). All of these disparate labeling problems are similar from an optimization point of view.

When labels are explicitly grouped in a hierarchy, the costs in the energy are naturally structured. In this work, we characterize a class of energies as having *hierarchical costs*. If an energy satisfies our "h-metric" and "h-subset" condi-

A. Delong (✉) · L. Gorelick · O. Veksler · Y. Boykov
Department of Computer Science, University of Western Ontario, London, ON, Canada N6A 5B7
e-mail: andrew.delong@gmail.com

L. Gorelick
e-mail: lena.gorelick@gmail.com

O. Veksler
e-mail: olga@csd.uwo.ca

Y. Boykov
e-mail: yuri@csd.uwo.ca