Int J Comput Vis (2012) 100:190–202
DOI 10.1007/s11263-011-0512-5

Motion Coherent Tracking Using Multi-label MRF Optimization

David Tsai · Matthew Flagg · Atsushi Nakazawa · James M. Rehg

Received: 26 December 2010 / Accepted: 1 December 2011 / Published online: 21 December 2011

Abstract We present a novel off-line algorithm for target segmentation and tracking in video. In our approach, video data is represented by a multi-label Markov Random Field model, and segmentation is accomplished by finding the minimum energy label assignment. We propose a novel energy formulation which incorporates both segmentation and motion estimation in a single framework. Our energy functions enforce motion coherence both within and across frames. We utilize state-of-the-art methods to efficiently optimize over a large number of discrete labels. In addition, we introduce a new ground-truth dataset, called Georgia Tech Segmentation and Tracking Dataset (GT-SegTrack), for the evaluation of segmentation accuracy in video tracking. We compare our method with several recent on-line tracking algorithms and provide quantitative and qualitative performance comparisons.

Electronic supplementary material The online version of this article (doi:10.1007/s11263-011-0512-5) contains supplementary material, which is available to authorized users.

D. Tsai (✉) · M. Flagg · J.M. Rehg
Center for Behavior Imaging and the Computational Perception Laboratory, School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA
e-mail: caihsiaoster@gatech.edu

M. Flagg
e-mail: mflagg@gmail.com

J.M. Rehg
e-mail: rehg@cc.gatech.edu

A. Nakazawa
Cybermedia Center, Osaka University, 1-32 Machikaneyama, Toyonaka, Osaka 560-0043, Japan
e-mail: nakazawa@cmc.osaka-u.ac.jp

Keywords Video object segmentation · Visual tracking · Markov random field · Motion coherence · Combinatoric optimization · Biotracking

1 Introduction

Recent work in visual target tracking has explored the interplay between state estimation and target segmentation (Bibby and Reid 2008; Chockalingam et al. 2009; Ren and Malik 2007). In the case of active contour trackers and level set methods, for example, the state model of an evolving contour corresponds to a segmentation of target pixels in each frame. One key distinction, however, between tracking and segmentation is that tracking systems are designed to operate automatically once the target has been identified, while systems for video object segmentation (Bai et al. 2009; Wang et al. 2005; Rother et al. 2004) are usually interactive, and incorporate guidance from the user throughout the analysis process. A second distinction is that tracking systems are often designed for on-line, real-time use, while segmentation systems can work off-line and operate at interactive speeds.

Several recent works have demonstrated excellent results for on-line tracking in real-time (Bibby and Reid 2008; Chockalingam et al. 2009). However, the quality of the segmentations produced by on-line trackers is in general not competitive with those produced by systems for interactive segmentation (Bai et al. 2009; Price et al. 2009; Li et al. 2005), even in cases where the user intervention is limited. One reason is that segmentation-based methods often adopt a global optimization method (e.g. graphcut) and explicitly search a large, fine-grained space of potential segmentations. In contrast, for tracking-based methods the space of possible segmentations is usually defined implicitly