Int J Comput Vis
DOI 10.1007/s11263-013-0666-4

Multi-Target Tracking by Online Learning a CRF Model of Appearance and Motion Patterns

Bo Yang · Ramakant Nevatia

Received: 21 December 2012 / Accepted: 7 October 2013

Abstract We introduce an online learning approach for multi-target tracking. Detection responses are gradually associated into tracklets in multiple levels to produce final tracks. Unlike most previous approaches which only focus on producing discriminative motion and appearance models for all targets, we further consider discriminative features for distinguishing difficult pairs of targets. The tracking problem is formulated using an online learned CRF model, and is transformed into an energy minimization problem. The energy functions include a set of unary functions that are based on motion and appearance models for discriminating all targets, as well as a set of pairwise functions that are based on models for differentiating corresponding pairs of tracklets. The online CRF approach is more powerful at distinguishing spatially close targets with similar appearances, as well as in tracking targets in presence of camera motions. An efficient algorithm is introduced for finding an association with low energy cost. We present results on four public data sets, and show significant improvements compared with several state-of-art methods.

Keywords Multi-target tracking · Online learned CRF · Appearance and motion patterns · Association based tracking

Electronic supplementary material The online version of this article (doi:10.1007/s11263-013-0666-4) contains supplementary material, which is available to authorized users.

B. Yang (✉) · R. Nevatia
Institute for Robotics and Intelligent Systems,
University of Southern California, Los Angeles, CA, USA
e-mail: yangbo@usc.edu; bo.yang02@gmail.com

R. Nevatia
e-mail: nevatia@usc.edu

1 Introduction

Automatic tracking of multiple targets is an important problem in computer vision, and is helpful for many real applications and high level analysis, e.g., event detection, crowd analysis, robotics. The aim of the task is to automatically locate concerned targets, give a unique id to each, and compute their trajectories. Due to significant improvements in object detection techniques in recent years, many researchers have proposed tracking methods that associate object detection responses into tracks, i.e., Association Based Tracking (ABT) methods (Perera et al. 2006; Xing et al. 2009; Andriyenko and Schindler 2011; Pirsiavash et al. 2011; Yang and Nevatia 2012a). They often apply an offline learned detector to each frame to provide detection responses, and then link these responses into tracklets, i.e., track fragments; tracklets are further associated into longer tracks in one or multiple steps. The probability of association of two tracklets is based on the motion smoothness and appearance similarity; a global solution with maximum probability is often found by Hungarian algorithm (Perera et al. 2006; Xing et al. 2009), MCMC (Yu et al. 2007; Song et al. 2010), or network flow analysis (Zhang et al. 2008; Pirsiavash et al. 2011), etc.

Association based approaches are powerful at dealing with extended occlusions between targets and the complexity is polynomial in the number of targets. To associate each target, motion and appearance information are often adopted to produce discriminative descriptors. Motion descriptors are often based on speed and distance between tracklet pairs, while appearance descriptors are often based on global or part based color histograms to distinguish different targets. However, how to better distinguish nearby targets and how to deal with motion in moving cameras remain key issues that limit the performance of ABT. In this paper, we propose

Published online: 23 October 2013