

# Discriminative Appearance Models for Pictorial Structures

Mykhaylo Andriluka · Stefan Roth · Bernt Schiele

Received: 31 July 2010 / Accepted: 16 September 2011 / Published online: 28 October 2011  
© Springer Science+Business Media, LLC 2011

**Abstract** In this paper we consider people detection and articulated pose estimation, two closely related and challenging problems in computer vision. Conceptually, both of these problems can be addressed within the pictorial structures framework (Felzenszwalb and Huttenlocher in *Int. J. Comput. Vis.* 61(1):55–79, 2005; Fischler and Elschlager in *IEEE Trans. Comput. C-22(1)*:67–92, 1973), even though previous approaches have not shown such generality. A principal difficulty for such a general approach is to model the appearance of body parts. The model has to be discriminative enough to enable reliable detection in cluttered scenes and general enough to capture highly variable appearance. Therefore, as the first important component of our approach, we propose a discriminative appearance model based on densely sampled local descriptors and AdaBoost classifiers. Secondly, we interpret the normalized margin of each classifier as likelihood in a generative model and compute marginal posteriors for each part using belief propagation. Thirdly, non-Gaussian relationships between parts are represented as Gaussians in the coordinate system of the joint between the parts. Additionally, in order to cope with shortcomings of tree-based pictorial structures models, we augment our model with additional repulsive factors in order to discourage overcounting of image evidence. We demon-

strate that the combination of these components within the pictorial structures framework results in a generic model that yields state-of-the-art performance for several datasets on a variety of tasks: people detection, upper body pose estimation, and full body pose estimation.

**Keywords** Object detection · People detection · Articulated pose estimation · Pictorial structures · Discriminative models

## 1 Introduction

People detection and articulated pose estimation are two challenging, long-standing problems in computer vision. Addressing these problems in real-world scenes is difficult because a successful approach should be both discriminative enough in order to distinguish people from the large amount of background clutter, and representative and flexible enough to capture the large variation in human appearance and poses. While numerous approaches have been proposed over the years, none of them has been demonstrated to be equally applicable for both people detection and pose estimation. Interestingly, quite a number of them (Andriluka et al. 2008; Ferrari et al. 2008; Ramanan 2007; Ramanan and Sminchisescu 2006; Zhang et al. 2006), while specializing on one of these tasks only, build on the same basic pictorial structures model (Felzenszwalb and Huttenlocher 2005; Fischler and Elschlager 1973). In this paper we show that given an appropriate representation for appearance and spatial components, the pictorial structures model obtains equal or even superior performance compared to many specialized approaches. This results in a generic model equally applicable for human detection and pose estimation, which enables detection of both upright people (i.e., pedestrians,

---

M. Andriluka (✉) · B. Schiele  
MPI Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken,  
Germany  
e-mail: andriluka@mpi-inf.mpg.de

B. Schiele  
e-mail: schiele@mpi-inf.mpg.de

S. Roth  
Department of Computer Science, TU Darmstadt,  
Fraunhoferstr. 5, 64283 Darmstadt, Germany  
e-mail: sroth@cs.tu-darmstadt.de