Int J Comput Vis (2012) 99:25–52
DOI 10.1007/s11263-012-0516-9

Fast Human Pose Detection Using Randomized Hierarchical Cascades of Rejectors

Grégory Rogez · Jonathan Rihan · Carlos Orrite-Uruñuela · Philip H.S. Torr

Received: 24 April 2011 / Accepted: 5 January 2012 / Published online: 31 January 2012

Abstract This paper addresses human detection and pose estimation from monocular images by formulating it as a classification problem. Our main contribution is a multi-class pose detector that uses the best components of state-of-the-art classifiers including hierarchical trees, cascades of rejectors as well as randomized forests. Given a database of images with corresponding human poses, we define a set of classes by discretizing camera viewpoint and pose space. A bottom-up approach is first followed to build a hierarchical tree by recursively clustering and merging the classes at each level. For each branch of this decision tree, we take advantage of the alignment of training images to build a list of potentially discriminative HOG (Histograms of Orientated Gradients) features. We then select the HOG blocks that show the best rejection performances. We finally grow an ensemble of cascades by randomly sampling one of these HOG-based rejectors at each branch of the tree. The resulting multi-class classifier is then used to scan images in a sliding window scheme. One of the properties of our algorithm is that the randomization can be applied on-line at no extra-cost, therefore classifying each window with a different ensemble of randomized cascades. Our approach, when compared to other pose classifiers, gives fast and efficient detection performances with both fixed and moving cameras. We present results using different publicly available training and testing data sets.

Keywords Human detection · Pose estimation · Cascade classifiers

1 Introduction

Full-body human pose analysis from monocular images constitutes one of the fundamental problems in Computer Vision as shown by the recent special issue of the journal (Sigal and Black 2010). It has a wide range of potential applications such as Human-Computer interfaces, video-games, video annotation/indexing or surveillance. Given an input image, an ideal system would be able to localize any humans present in the scene and recover their poses. The two stages, known as *human detection* and *human pose estimation*, are usually considered separately. There is an extensive literature on both *detection* (Viola et al. 2005; Wu and Nevatia 2005; Dalal and Triggs 2005; Zhu et al. 2006; Gavrila 2007; Sabzmeydani and Mori 2007) and *pose estimation* (Shakhnarovich et al. 2003; Agarwal and Triggs 2006; Mori and Malik 2006; Thayananthan et al. 2006; Bissacco et al. 2007; Rogez et al. 2008a; Jaeggli et al. 2009; Elgammal and Lee 2009; Lee and Elgammal 2010) but relatively few papers consider the two stages together (Dimitrijevic et al. 2006; Bissacco et al. 2006; Sminchisescu et al. 2006; Okada and Soatto 2008; Bourdev and Malik 2009). Most algorithms for pose estimation assume that the human

Part of this work was conducted while the first author was a research fellow at Oxford Brookes University. This work was partly supported by the EPSRC grant GR/T21790/01(P) and by Sony Entertainment Europe (SCEE). G. Rogez and C. Orrite would like to acknowledge support provided by: "Departamento de Ciencia, Tecnología y Universidad del Gobierno de Aragón", "Fondo Social Europeo" and "Ministerio de Ciencia e Innovación (TIN2010-20177)". Prof. Torr is in receipt of a Royal Society Wolfson Research Merit Award.

G. Rogez (✉) · C. Orrite-Uruñuela
Computer Vision Lab, Aragon Institute for Engineering Research (I3A), University of Zaragoza, Zaragoza, Spain
e-mail: grogez@unizar.es

J. Rihan · P.H.S. Torr
Department of Computing, Oxford Brookes University, Wheatley Campus, Oxford OX33 1HX, UK