Int J Comput Vis
DOI 10.1007/s11263-013-0689-x

Domain Adaptation for Structured Regression

Makoto Yamada · Leonid Sigal · Yi Chang

Received: 15 March 2013 / Accepted: 30 November 2013

Abstract Discriminative regression models have proved effective for many vision applications (here we focus on 3D full-body and head pose estimation from image and depth data). However, *dataset bias* is common and is able to significantly degrade the performance of a trained model on target test sets. As we show, covariate shift, a form of unsupervised domain adaptation (USDA), can be used to address certain biases in this setting, but is unable to deal with more severe structural biases in the data. We propose an effective and efficient semi-supervised domain adaptation (SSDA) approach for addressing such more severe biases in the data. Proposed SSDA is a generalization of USDA, that is able to effectively leverage labeled data in the target domain when available. Our method amounts to projecting input features into a higher dimensional space (by construction well suited for domain adaptation) and estimating weights for the training samples based on the ratio of test and train marginals in that space. The resulting augmented weighted samples can then be used to learn a model of choice, alleviating the problems of bias in the data; as an example, we introduce SSDA twin Gaussian process regression (SSDA-TGP) model. With this model we also address the issue of *data sharing*, where we are able to leverage samples from certain activities (e.g., walking, jogging) to improve predictive performance on very different activities (e.g., boxing). In addition, we analyze the relationship between domain similarity and effectiveness of proposed USDA versus SSDA methods. Moreover, we propose a computationally efficient alternative to TGP (Bo and Sminchisescu 2010), and it's variants, called the direct TGP. We show that our model outperforms a number of baselines, on two public datasets: HumanEva and ETH Face Pose Range Image Dataset. We can also achieve 8–15 times speedup in computation time, over the traditional formulation of TGP, using the proposed direct formulation, with little to no loss in performance.

Keywords 3D pose estimation · Semi-supervised domain adaptation · Covariate shift adaptation

M. Yamada (✉) · Y. Chang
Yahoo Labs, 701 1st Ave., Sunnyvale, CA 94089, USA
e-mail: makotoy@yahoo-inc.com

Y. Chang
e-mail: yichang@yahoo-inc.com

L. Sigal
Disney Research Pittsburgh, 4720 Forbes Ave.,
Pittsburgh, PA 15213, USA
e-mail: lsigal@disneyresearch.com

1 Introduction

Many problems in computer vision can be expressed in the form of discriminative (structured) predictions of real-valued multivariate output, $\boldsymbol{y} \in \mathbb{R}^{d_{\mathrm{y}}}$, from a high-dimensional multivariate input, $\boldsymbol{x} \in \mathbb{R}^{d_{\mathrm{x}}}$. A success of such methods in 3D full-body pose estimation is evident from recent results that use Microsoft Kinect sensor (Girshick et al. 2011; Sun et al. 2012); such discriminative methods have also proved effective for other problems, including image-based 3D pose (Bo and Sminchisescu 2010; Kanaujia et al. 2007; Shakhnarovich et al. 2003; Sminchisescu et al. 2006; Urtasun and Darrell 2008) head pose (Fanelli et al. 2011) and body shape (Chen et al. 2011; Sigal et al. 2007) estimation. The typical goal of discriminative regression methods is to learn a direct (and sometimes multi-modal) mapping, $\boldsymbol{f} : \mathbb{R}^{d_{\mathrm{x}}} \rightarrow \mathbb{R}^{d_{\mathrm{y}}}$, from features (e.g., computed from image or depth data) to pose (e.g., 3D position and orientation of the head, or full 3D pose of the body encoded by joint positions or joint angles). Despite success and large body of work, most approaches

Published online: 13 December 2013