Learning Radon-Nykodim Derivatives for Unsupervised, Real-time Object Detection

Ian Fasel
Department of Cognitive Science
University of California San Diego
La Jolla, CA 92093-0515
ianfasel@mplab.ucsd.edu

Javier R. Movellan
Institute for Neural Computation
University of California San Diego
La Jolla, CA 92093-0523
movellan@mplab.ucsd.edu

We propose a new method for using discriminative approaches within a generative framework for weakly-supervised learning. We focus on the task of object detection, i.e., provided an image, our goal is to locate an object of interest within the image. Rather than require that the exact location of the object is provided in each training example, we instead require only a label indicating whether or not the object of interest is present somewhere in the image. This setting for learning is identical to the setting of Weber, Welling, Perona (2000), however because at run-time we wish to identify the precise location and size of the object in new images, not just it's presence or absence, and we wish to do this in real-time (at least 15fps) on commodity hardware, we require a significantly different approach.

We use an image generation model in which foreground objects, rendered according to an unknown foreground distribution, occlude a background scene, rendered according to an unknown background distribution. To perform optimal inference under this model, one must consider every region of the image (with non-zero prior) and estimate the maximum likelihood *ratios* of the pixels in the region. This means that rather than fully learning the foreground and background distributions, it is sufficient to learn a model just of the likelihood ratios. Leveraging this fact, we use a log-linear model, or a "random field", to learn the Radon-Nykodim derivative of the foreground model with respect to the background model. This formulation of the problem allows us to apply a functional gradient descent approach to perform efficient stage-wise additive feature-selection, resulting in an algorithm very similar to AdaBoost, with a model capable of performing real-time object detection using Haar wavelet-like features (*ala* Viola & Jones, 2001).

Early results show that this technique is competitive with AdaBoost when the data is fully labeled (i.e., every example image region is labeled as 'object' vs. 'background'), however unlike AdaBoost this method can also efficiently learn when the examples are not fully labeled (i.e., images are only labeled as 'one of the regions contains the object' vs. 'none of the regions contains the object'). We discuss the algorithm in detail, including the use of second-order approximations to improve learning, training-time, as well as challenges, such as how to restrict learning to valid Radon-Nykodim derivatives.