Object Class Recognition Using Multiple Instance Learning with Objectness as Location Prior

-- Shadab Khan
COSC 174, Dartmouth College
Project Milestone Report

Introduction

Object class recognition in a given image is a difficult problem. To classify a test image as hit or match for a particular object class requires the abstraction of an object, developed using real world images. This is the where the difficulty lies - since the images of an object class can have huge variations in terms of illumination, angle of view, scale, rotation, color, location, type etc (intra-class variations). For example, the different photographs of the object class 'car' from PASCAL VOC 2007 [1] dataset shown below differ from each other in aspects mentioned previously.

Fig. 1 Images showing object class 'car' from PASCAL VOC 2007 dataset. Note the variations in the illumination, scale, rotation, angle of view, occlusion, manufacturer and model etc.

This necessitates use of complex algorithms to develop as robust abstraction as possible within computational budget to handle the variations while still producing useable classification performance. The complete process of object class recognition can be understood as a three step process involving : 1) Representation of the object, 2) Learning an abstraction or a model using set of training images which are represented by method developed in previous step, 3) Testing the object class recognition. A discussion of proposed approach for these steps follows.

Representation of the object

To represent each image, we need image-features which are not only highly discriminative, but are also robust to the variations. To decide which features to use, I did a survey of features which are typically used for object class recognition and shortlisted the following features.

1. Bag of Visual Words : A commonly used technique to represent image as a histogram of visual words. Introduced by Csurka et al. in 2004 [2], this has grown up to be a very popular method of image representation. This works by first finding a number of interest points in an image - this

step is often called ‘detection’. Next, the region around interest points is windowed and abstracted - this step is often called ‘description’. These descriptors are then quantized into visual words, often using k-means clustering with Mahalanobis or Euclidean distance [3]. The feature vector then contains the number of occurrence of each visual word in an image.

2. GIST : Developed by Torralba et al. [4], this algorithm computes the histograms of gradient orientations which are localized at test point. It roughly abstracts the spatial arrangement of image structures and has been proven to work well for describing the general appearance of scene.
3. Histogram of Oriented Gradients (HOG): Developed by Dalal and Triggs [6], this is a well-known feature that has shows its promise in state-of-the-art part-based object detector [5].

Although several more features can be computed, these features make for a good choice and have been shown to work well for object-class recognition problem by L Torressani et al. [7]. For the Bag of Visual Words feature, SURF (Speeded Up Robust Features) will be used, which is a fast and scale-invariant, rotation-invariant detector and descriptor [8].

Learning a Model

I have looked at some of the existing Multiple Instance Learning algorithms, however, a decision is yet to be made of the MIL algorithms that will be tried.

Progress so far

I have the code in place to compute the features. To create of a bag of images from training set, I am using Objectness measure [9] to sample windows of an image from training set. This code is available too. In the time ahead., I will implement the MIL algorithms to test the classification performance.

References

[1] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 Results", Web: http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html

[2] G. Csurka, C R Dance, L Fan, J Williamowski, C Bray, “Visual Categorization with Bags of Keypoints”, in *Proceeding of ECCV*, 2004.

[3] **BOOK:** Computer Vision: Algorithms and Applications, Richard Szeliski, Springer Publications, 2010.

[4] Aude Oliva, Antonio Torralba, “Modeling the shape of the scene: a holistic representation of the spatial envelope”, *International Journal of Computer Vision*, Vol. 42(3): 145-175, 2001.

[5] P Felzenszwalb, R B Girshick, D. McAllester and D Ramanan, “Object detection with discriminatively trained part based models” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[6] N Dalal and B Triggs “Histogram of Oriented Gradients for Human Detection”, in *Proceeding of CVPR*, 2005 .

[7] Alessandro Bergamo, Lorenzo Torresani, Andrew Fitzgibbon “PiCoDes: Learning a Compact Code for Novel-Category Recognition” in *Proceeding of Neural Information Processing Systems (NIPS)*, 2011.

[8] H Bay, A Ess, T Tuytelaars, L V Gool, “SURF: Speeded up robust features”, *Computer Vision and Image Understanding*, pp. 346-350, 2008.

[9] V Ferrari et al. “What is an Object?”, CVPR, 2010.