Paper Critique - A Bayesian Hierarchical Model for Learning Natural Scene Categories

Karim Ali

karim@uwaterloo.ca

David R. Cheriton School of Computer Science

University of Waterloo

1 Overview

The paper starts off by explaining why supervised learning learn and recognize natural scene categories is no longer a desirable approach. The main reason is that supervised learning in that case depends on the manual annotation of intermediate properties of natural scenes to be able to come up with a categorization scheme for the images under analysis. Of course, manual image annotation is both a tedious and an expensive task to do. In addition, exper-defined labels are some what arbitrary and possibly sub-optimal.

The main contribution of the paper is providing a Bayesian hierarchical model to learn and recognize natural scene categories without supervision. The authors extended the notion of a **texton** (i.e. codeword) from texture and material literature and apply it to images. The algorithm starts by giving a label to the image (manually). The image is then classified based on how many codewords from a specific category (the set of categories depend on the category label given to the image) are found in it. The process goes on until all training images are processed. Then, the process of recognizing natural scenes from test images starts (in a similar fashion, only that it's the category label that we're after).

* What's the take home message?

2 Significance and Originality

The paper can be considered as an extension/application of the idea of using **textons**, in the context of natural scene categorization. The novelty of the approach is that it reduces the dependence on human effort compared to previous methods that follow a supervised learning approach.

The citation count of the paper (around 500) also suggest that the paper is heavily referred to in the field. This certainly shows how significant the contributions of the paper are.

3 Soundness

Although the authors claim that the learning process is an unsupervised/automatic process, they still depend on human efforts to assign the category labels for the training images. This step is crucial for the algorithm, otherwise the algorithm will not be able to decide which set of codewords it should mark as **relevant**. An improvement might be to automate the process of assigning category labels by following the same logic of the algorithm used to assign a test image to a category. In other words, given an image, it will be divided into patches (

4   Evaluation

* Are the ideas/algorithms empirically or theoretically evaluated?
   * If some assumptions are made, are they realistic? (although I did not see in their paper referenecs to back up this claim)
   * Is scalability demonstrated?

5   Related Work

* Is the paper properly situated with respect to related work? * Is there a brief survey of related work? * Do the authors explain similarities and differences with previous work?

6   Readability

* Is the paper well structured? * Does the abstract properly and accurately summarize the paper? * Do the introduction and conclusion clearly explain the contributions and the take home message? * Is the flow of ideas easy to follow? * Is the paper well written? * Are all the technical terms and abreviations explained? * Are there important grammatical errors? * Are there a lot of typos? just one in page 5

References

[1] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005.*, vol. 2. IEEE, 2005, pp. 524–531.