Abstract

Daniel Matthew Steinberg
The University of Sydney

Doctor of Philosophy
July 2013

An Unsupervised Approach to Modelling Visual Data

With the advent of cheap, high fidelity, digital imaging systems it is now easy to create huge collections of digital images. Subsequently, the computer vision community has seen an explosion of research in classifying these images into scenes, recognising objects within images, propagating user tags to new images, and even attempts at whole image "understanding".

Most of this research uses *supervised* or *semi-supervised* algorithms, which rely upon some form of human generated "ground-truth". For very large scientific datasets with many classes, producing the ground-truth data can represent a substantial, and potentially expensive, human effort. In these situations there is scope for the use of unsupervised approaches that can model collections of images and automatically summarise their content. The primary motivation for this thesis comes from the problem of labelling large visual datasets of the seafloor obtained by an autonomous underwater vehicle (AUV) for ecological analysis. It is expensive to label this data, as taxonomical experts for the specific region are required. Quick, approximate summaries of quasi-habitats and objects within images can be generated by unsupervised methods "for free". These can be used to focus the efforts of experts, and inform decisions on additional sampling. These techniques are equally applicable to large photo albums and collections, such as the millions of images hosted on sites like *Flickr*, where image annotations may be incorrect or absent entirely.

The contributions in this thesis arise from modelling this visual data in entirely unsupervised ways to obtain comprehensive visual summaries for subsequent expert annotation. Firstly, popular unsupervised image feature learning approaches are adapted to work with large datasets and unsupervised clustering algorithms. Next, using Bayesian models the performance of rudimentary scene clustering is boosted by sharing clusters between multiple related datasets, such as photo albums or AUV surveys. Then these Bayesian scene clustering models are extended to simultaneously cluster sub-image super-pixels, or segments, to form unsupervised notions of "objects" within scenes. The frequency distribution of these objects within scenes is used as the scene descriptor ("bag-of-segments") for simultaneous scene clustering. This model also takes advantage of multiple related datasets, and its various properties are shown to enhance clustering through the use of contextual information inherent within the data. Finally, this simultaneous clustering model is extended to make use of whole image descriptors, which encode rudimentary spatial information, as well as object frequency distributions to describe scenes. This is achieved by unifying the previously presented Bayesian clustering models, and in so doing rectifies some of their weaknesses and limitations. Hence, the final contribution of this thesis is a practical *unsupervised* algorithm for modelling images from the super-pixel to album levels, and is applicable to large datasets.