EFFICIENT AND COMPACT CODEBOOK DESIGN FOR SCENE ANALYSIS AND OBJECT LOCALISATION

B. Mayurathan

Postgraduate Institute of Science, University of Peradeniya, Sri Lanka

Computer vision is the science of endowing computers or other machines with vision, or the ability to see. The ultimate goal of computer vision is to model and replicate human vision using computer software and hardware of different levels. It combines knowledge in computer science, electrical engineering, mathematics, biology and cognitive science in order to understand and simulate the operation of the human vision system. Visual scene recognition and object localisation are important areas of study in computer vision. Several applications including the design of biometric systems for security, automatic number plate recognition and content-based image retrieval motivate intense research interest in this subject. Also, detecting visual objects in images is a very important component in computer vision systems such as image retrieval, scene understanding and surveillance system; however, it is still an open problem because the intraclass variations make generic detection very complicated, requiring various types of pre-processing steps. Further, research into artificial vision systems helps in developing a better understanding of how the human visual system might store, retrieve and recognize thousands of classes of objects so efficiently.

A particularly successful computational approach to visual scene recognition starts from a bag-of-words representation of low level image features. When low level features, usually relating to sharp gradient changes, are extracted from images, the cardinality (*i.e.* number of features extracted) can vary due to background noise, presence of other objects in the scene, camera angle and lighting etc. This poses a fundamental problem in applying standard machine learning techniques to visual object recognition.

The bag-of-features approach gives a way of mapping any image from this variable numbers of features to a fixed dimensional space. This is achieved by vector quantization of the extracted low level features against a codebook, and representing the image as a histogram of how many times each of the code-words was seen in the extracted set of features. Hence the construction of codebook, usually achieved by k-means clustering of a large number of low level features, is an important step in designing visual scene recognition and localisation systems. It has been noted previously that the clustering process is both a computational and performance bottleneck, and scaling it to large scene recognition tasks is a major challenge.

In this dissertation, we address the codebook design (clustering) problem by means of sequential one-pass algorithms, which offer computationally efficient results because each example is processed only once. Examples that are redundant with respect to our objective are removed from further consideration. The starting points for our research are the Resource Allocation Network, the Sequential Input Space Partitioning algorithm for training multi-layer perceptron classifiers and the more recent attempt at adapting these ideas to clustering – the Resource Allocating Codebook approach. We build on these ideas and develop two novel algorithms, the Sequential Input Space Carving (SISC) approach and an enhanced version of Sequential Input Space Carving (which we refer to as SISC^{++}).

We show empirically, using standard benchmark datasets, that the algorithms we propose have the property of delivering computationally efficient and compact codebooks, while retaining the discriminative power that other authors (and ourselves) have achieved using codebooks designed by the k-means algorithm. The benchmark datasets we have used for these demonstrations cover a wide spectrum, and include visual scene classification, human action classification and texture classification. Somewhat related algorithms in the literature are DBSCAN and Mean Shift. We argue that the approach we propose offers substantial computational advantage over these methods.

Additionally, this thesis also demonstrates a method to predict the bounding box of an object. Here, we proposed a binary histogram method which is based on foreground and background of image features to detect the bounding box of an object. Given an image and an object category, our goal is to predict a bounding box indicating where the object of interest is as well as inferring the supporting pixels for the object.

Index terms: Scene recognition, object detection, benchmark datasets, visual vocabularies, bag-of-words, Sequential Input Space Carving (SISC), Enhanced Sequential Input Space Carving ($SISC^{++}$).