

Creating Compact and Discriminative Visual Vocabularies Using Visual Bits

T. Kirishanthy and A. Ramanan

Department of Computer Science, Faculty of Science, University of Jaffna, Sri Lanka
kirishanthy.t@gmail.com, a.ramanan@jfn.ac.lk

Abstract

In a patch-based object recognition system the key role of a visual vocabulary is to provide a way to map the low-level features into a fixed-length vector in histogram space to which standard classifiers can be directly applied. The discriminative power of such a visual vocabulary determines the quality of the vocabulary model, whereas the size of the vocabulary controls the complexity of the model. A compact visual vocabulary provides a lower-dimensional representation whereas a large-sized vocabulary may overfit to the distribution of visual words in an image and lead to heavy computational load. The generic framework of a bag-of-features approach follows a standard routine extracting local image descriptors and clustering with a user-designated number of clusters. The problem with this routine lies in that constructing a vocabulary for each single dataset is not efficient. Usually the construction of a vocabulary is achieved by cluster analysis using K-means algorithm. However, one of its drawbacks is the choice of a suitable value for K which determines the size of a visual vocabulary. The choice of the size of a vocabulary should be balanced between the recognition rate and computational needs. In this paper we propose a two-staged approach to map an initial high-dimensional vocabulary into a compact vocabulary while maintaining its discriminative power. Using an initial larger vocabulary we first represent the training images using a coding scheme that maps the importance of each visual word within an image as visual bits. These set of visual bits of images then form a sparse representation of every visual word with respect to the set of category-specific training images that is used for the compression. We have tested our vocabulary compression technique on four computer vision tasks: (i) Xerox7 (ii) PASCAL VOC Challenge 2007 (iii) UIUC texture and (iv) MPEG7 CE Shape-1 Part B Silhouette image datasets. Testing results show that the proposed method slightly outperforms vocabularies learnt by K-means by achieving just half the size of initial vocabulary. Our compression technique could help to optimize larger vocabularies to fewer visual words with stable performance.

Author Keywords

Bag-of-features; Compact Visual Vocabulary; Object recognition; Visual bit representation