28 Jan 2014

Object Categorization by Learned Universal Visual Dictionary (UVD)

Visual Dictionaries generally have an “optimal” size at which they give best categorization accuracy. Hence, dictionary sizes are manually adjusted in the beginning. This paper proposes a supervised approach to learn a compact and discriminative visual dictionary from an initial large dictionary. An ideal dictionary would make histograms of same class to be similar (less intra class variation) and different classes to be different (more inter class variation). In this approach, the compact dictionary is learnt by sequentially merging two bins which do not help in discriminating between classes. This is done unto the point when all possible merging lead to reducing the discriminating power. Performing object categorization using the learned dictionary gives similar accuracies as the initial dictionary but is much faster because of the small size. Also, the proposed learning algorithm is very efficient.

Once the dictionary is learned, object classification can either be done through nearest neighbour approach (non-parametric) or by explicitly modelling the classes through Gaussian models (parametric). The authors experimented with multiple datasets and showed that using parametric or non-parametric methods with the UVD performs nearly similar to using nearest neighbour with the initial large dictionary. The effect of varying initial dictionary size on the UVD performance is shown. Also, the class discrimination ratio was found to increase with the learned dictionary.

The approach is very beneficial in terms of classification efficiency. Handling large dictionaries especially in case of non-parametric methods is computationally expensive. Also, the authors have shown that performance increases with the size of initial dictionary, so we no longer have to worry about an optimal dictionary size. The approach could have been shown more general if dictionaries of different types were considered: for example using local descriptors other than textons like SIFT to create the dictionary. Also, it would have been interesting to see the performance of the UVD on other tasks like image retrieval etc. Similar techniques can be very useful in reducing the size of extremely large n-gram dictionaries, where the bins are not individual visual words but combinations of words.

A Performance Evaluation of Local Descriptors

This paper compares the performance of various interest point descriptors namely SIFT, steerable filters, differential invariants, moment invariants, complex filters, cross correlation (simplest descriptor). The descriptors were combined with various interest point detectors like Harris, Harris-Laplace detectors etc, some of which are scale invariant while others are rotation, affine transformation invariant. The performance of the descriptors was measured in terms of correct point matches found between transformed images of same scene. Various image transformations like scale, rotation, illumination and affine were considered. The effect of various errors in interest point detectors like errors in point localization, scale estimation error, orientation estimation was also studied.

The paper identified SIFT as the best performing descriptor over all the transformations. Also, the ranking of the descriptors did not depend on the underlying interest point detector or the image transformations. As expected, the overall performance of the descriptors degraded with difficult transformations like affine and scale-rotation.