Image Categorization
Using Spatial Matching and Indexing

Agni Delvinioti

October 24, 2012

Abstract

In the framework of this Diploma thesis we introduce a new image categorization method, which integrates spatial matching and indexing in the classification process. Spatial matching is based on *Hough pyramid matching* (HPM); indexing is based on an inverted file structure as in image retrieval; and classification is carried out with a multiclass *support vector machine* (SVM) classifier.

We use HPM as an image similarity measure and we show that under reasonable assumptions it is a Mercer kernel. We do so by explicitly expressing it as an inner product in a high dimensional space where images lie given an appropriate quantized representation of their local features and descriptors. We then use this kernel for SVM training instead of a linear kernel, which is a typical choice under the *bag of words* (BoW) model. It is the first time that a kernel function takes spatial configuration into account while being invariant to translation, scale and rotation. In most cases, artificial perturbations are the only way to achieve geometric invariance, with an exponential increase of training time.

We train one binary SVM classifier for each category following an one-versus-all strategy and then combine individual classifiers into one multiclass classifier. Comparing to nearest-neighbor classifier using e.g. image retrieval methods, we exploit the sparse representation of SVMs: at classification time, the query image is matched via HPM against the chosen support vectors only. However, matching need not be exhaustive. Support vectors are indexed into an inverted file, and HPM may be applied only to a small subset that is top-ranking according to any scalar similarity measure, e.g. based on BoW. The method therefore easily applies to large scale classification, while training for unseen classes does not require re-training for existing ones.

Due to the nature of local features and their use in invariant matching, the method is most appropriate for specific object recognition. We apply it to landmark recognition, conducting experiments on our own dataset, constructed from the World cities dataset via a semi-automatic process that combines visual and geographical clustering. We compare to a baseline classifier using a BoW representation and achieve more than a twofold increase in accuracy on experiments of up to 68 landmarks.

Keywords

Image categorization, kernel learning, spatial matching, indexing, image retrieval, landmark recognition