Abstract

In this work the problems of specific object and image retrieval including the more challenging sub-image are studied. Given a query image of a specific object a retrieval engine returns relevant images of the same object from a database. The thesis focuses on the bag-of-words approach which is one of the most effective content-based approach especially when the specific object covers only a part of the picture, can be occluded or only partially visible. The thesis improves a number of components of the standard bag-of-words retrieval approach.

A novel similarity measure for bag-of-words type large scale image retrieval is presented. The similarity function is learned in an unsupervised manner, requires no extra space over the standard bag-of-words method and is more discriminative than both L2-based soft assignment and Hamming embedding. The novel similarity function achieves mean average precision that is superior to any result published in the literature on the standard datasets and protocols.

We study the effect of a fine quantization and very large vocabularies (up to 64 million words) and show that the performance of specific object retrieval increases with the size of the vocabulary. This observation is in contradiction with previously published results. We further demonstrate that the large vocabularies increase the speed of the tf-idf scoring step.

All state-of-the-art image retrieval results in the literature have been achieved by methods that include a query expansion which brings a significant boost in performance. We introduce three modifications to automatic query expansion: (i) a method capable of preventing *query expansion failure* caused by the presence of confusers, (ii) an improved spatial verification and re-ranking step that incrementally builds a statistical model of the query object and (iii) we learn relevant spatial context to boost retrieval performance.

All three improvements of query expansion were evaluated on standard Paris and Oxford datasets and state-of-the-art results were achieved.

Finally, novel problems for image retrieval are formulated. It is shown that the classical ranking of images based on similarity addresses only one of possible user requirements. Instead of searching for the most similar images, the novel retrieval methods zoom-in and zoom-out answer the *"What is this?"* and *"Where is this?"* questions.

In addition, two other task are formulated: (i) given a query and a large image dataset, for every pixel location in the query, find an image with maximum resolution and (ii) return the frequency with which a pixel appears in the dataset.

The zoom-in and zoom-out required the development of two novel techniques: the hierarchical query expansion method and a geometric consistency verification step that is sufficiently robust to prevent a topic drift within a zooming search. Experiments show that the proposed methods find surprisingly fine details on the tested landmarks, even those that are hardly noticeable for humans.