



Practical computer vision examples: indexing and categorization

Andrea Vedaldi
University of Oxford, United Kingdom

Abstract

This course is a practical introduction to a number of fundamental techniques in image understanding. In a series of guided experiments, students are asked to explore how such ideas can be implemented in software by using MATLAB and open source libraries.

The course is composed of two sessions. The first session explores instance-level recognition, which is the problem of matching in images a specific object or scene. Examples include recognizing a specific building, such as Notre Dame, or a specific painting, such as 'Starry Night' by Van Gogh. The object is recognized despite changes in scale, camera viewpoint, illumination conditions and partial occlusion. An important application is image retrieval - starting from an image of an object of interest (the query), search through an image dataset to obtain (or retrieve) those images that contain the target object. The hands-on experience includes: (i) using SIFT features to obtain sparse matches between two images; (ii) using affine co-variant detectors to handle changes in viewpoint; (iii) vector quantizing the SIFT descriptors into visual words to enable large scale retrieval; and (iv) constructing and using an image retrieval system to identify objects.

The second session investigates category-level recognition, i.e. the problem of recognizing in images a type of objects or scenes, such as animals, people, or mountains. Objects are recognised despite changes in viewing condition as well as the intrinsic variability of the object class (no two mountains look exactly the same). As for object-level recognition, an important application is image retrieval - searching through an image dataset to find the images with a particular visual content. The hands-on experience includes: (i) training a visual classifier for five different image classes (airplanes, motorbikes, people, horses and cars); (ii) assessing the performance of the classifier by computing a precision-recall curve; (iii) varying the visual representation used for the feature vector, and the feature map used for the classifier; and (iv) obtaining training data for new classifiers using Bing image search.

Keywords: Hands-on experience with SIFT feature extraction and matching, image indexing and retrieval, bag-of-visual words, learning of SVMs, and image categorization.