平成24年度 メディア科学専攻博士論文要旨

大西 研究室	氏　　名	Shuang Bai
博士論文題目	A Study of Informative Patch Extraction and Patch-level Context Exploration for Image Categorization	

This thesis deals with two problems for image categorization under the bag of visual words framework. The first problem is regarding image patch extraction. Until now, the commonly used image patch extraction methods could be divided into two main branches. The first branch is based on dense sampling or interest point detectors. These methods are not able to extract informative image patches for the task of image categorization. The other branch is through exhaustive sampling together with a feature selection process. It has high costs in computation and memory usage. Therefore, to cope with this problem, we focus on designing an effective image patch extraction approach. The second problem is concerned with patch-level context exploration. In order to create discriminative image representations, we try to explore patch-level context to incorporate relationships among patches. Although some work on patch-level context already exists, the combination of patches of interest and their context patches is fixed in such works. Variations of the relationships among these patches are not taken into consideration. In this thesis, we propose an approach in which the patch of interest is flexibly combined with its context patches.

First, we present the patch extraction method. In the proposed method, both bottom-up processing and top-down processing are utilized for evaluating image regions with respect to their informativeness for image categorization. Given an input image, the bottom-up branch evaluates it through processing its low-level image information, including interest point detection, interest point distribution mode localization and Gaussian mixture model modelling. On the other hand, in the top-down branch, images are evaluated by investigating the statistical property of the training image set. During this procedure, we collect local image features from training images and apply vector quantization to them to generate a cluster list. Then the mean vector, covariance matrix and entropy of each cluster are calculated and recorded. Thereafter, corresponding to each kind of processing, a separate saliency map is created; these are then fused with each other in a weighted manner. Finally, we perform image patch extraction based on the fused saliency map.

Next, we propose the patch-level context exploration method. In order to solve the ambiguity problem existing in the patch encoding process, a patch of interest is associated with other patches, such as patches extracted from the same sample point but of different scales or patches in its neighbourhood. When creating image representations, the associated patches are taken as a patch set and dealt with together, so the encoding of the patch of interest depends on not only itself but also its context patches. Furthermore, in order to explore the information in an input image extensively, we propose to utilize different kinds of context strategies and create a separate image representation corresponding to each context strategy. Consequently, we obtain three image representations for the input image, i.e. image representation without context, image representation with different scale patch context and image representation with nearby patch context. In the image classification stage, these representations are fused in a probabilistic way to predicate the input image label.

Finally, for each proposed method, we conducted extensive experiments to assess its effectiveness. We evaluated the patch extraction method using an object category dataset and a scene category dataset, and evaluated the patch-level context exploration method using three scene category datasets. We compared the proposed method to previous research. The results we obtained have demonstrated the effectiveness of the proposed methods.