Sparse Based Image Classification with Different Keypoints Descriptors

Yuanyuan Zuo⋆ and Bo Zhang

State Key Laboratory of Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science and Technology,
Tsinghua University, Beijing, 100084, China

Abstract. In this paper, we apply the sparse representation based algorithm to the problem of generic image classification. Keypoints with different descriptors are used as the bases of the training matrix and test samples. A learning algorithm is also presented to select the most important keypoints as the bases of the training matrix. Experiments have been done on 25 object categories selected from Caltech101 dataset, with salient region detector and different descriptors. The results show that keypoints with histogram of oriented gradients descriptor can achieve good performance on image categories which have distinctive patterns detected as keypoints. Furthermore, the base learning algorithm is useful for improving the performance while reducing the computational complexity.

Keywords: Image classification; sparse representation; keypoints.

1 Introduction

The task of image classification involves two important issues. One is image representation, the other is classification algorithm.

Recently, keypoints-based image features are getting more and more attention in the computer vision area. Keypoints, also known as interest points or salient regions, refer to local image patches which contain rich information, have some kind of saliency and can be stably detected under a certain degree of variations. Extraction of keypoints-based image feature usually includes two steps. First, keypoints detectors are used to find keypoints automatically. Second, keypoints descriptors are used to represent keypoints features. Ref. [1] and [2] gave a performance evaluation among several different keypoints detectors and descriptors respectively.

Corresponding to the different kinds of image representation, many classification algorithms have been proposed, which can be divided into two classes.

* The work is supported by the National Natural Science Foundation of China under Grant Nos. 90820305; the National Basic Research Program (973 Program) of China under Grant Nos. 2007CB311003.
One is generative models such as constellation model [3]. The other is discriminative models, such as support vector machines (SVM), which have been proved to be effective for object classification in [4-6].

Recently, sparse coding has been used for the learning of the codebook and image representation [7]. Wright et al. proposed sparse representation based classification [8] for the recognition of human faces. Although good performances have been achieved with the algorithm, the image database is strictly confined to human frontal faces with only illumination and slight expression changes.

In [9], we have applied the sparse representation based classification algorithm to the problem of generic image classification with a certain degree of background clutter, scale, translation, and rotation variations within the same image class. Bag of visual words features are used in the experiments. Comparable experimental results have been obtained with SVM classifiers under different size of vocabulary and numbers of training images.

However, bag of visual words features quantize local features into a given size of codebook and reflect the distribution of visual words detected on an image. Original local features of keypoints are omitted, which may contain distinctive patterns and be very helpful for recognizing some classes of images. In this paper, we propose local features for image classification based on sparse representation (Local-SRC). A base learning algorithm is also presented in order to select the most important keypoints as the bases of the training matrix.

The remainder of this paper is organized as follows. In section 2, the extraction method of keypoints is described. Section 3 gives a detailed description of the Local-SRC algorithm. A base learning algorithm is presented in section 4, followed by experiments and conclusions in section 5 and 6.

2 Image Keypoints Extraction

The keypoints extraction method includes the following two steps.

1) Keypoints detector. Salient region detector [10] proposed by Kadir et al. is one of the most widely used keypoints detectors. This detector selects regions which exhibit unpredictability both in the local attributes space and scale space. Unpredictability of image regions is measured by Shannon entropy of local image attributes, such as the pixel gray value. A value of saliency is computed for each region and the salient regions are sorted by the value of saliency. The amount of regions which are detected in one image usually varies from dozens to hundreds.

2) Keypoints descriptor. Histograms of Oriented Gradients (HOG) descriptor [11] is used to describe the feature for each keypoint. It computes gradients for every pixel in the keypoint local patch. The orientation of gradients (unsigned $0^\circ - 180^\circ$, or signed $0^\circ - 360^\circ$) is quantized to a certain number of bins. Local patches can be divided into different size of blocks, on which HOG features are computed. Ref. [11] experimented on different block sizes and normalization schemes. The results show that $2 \times 2$ blocks and $l^2$-norm perform well.