Chapter 2
Sparse Based Image Classification With Bag-of-Visual-Words Representations

Yuanyuan Zuo  
Tsinghua University, China

Bo Zhang  
Tsinghua University, China

ABSTRACT
The sparse representation based classification algorithm has been used to solve the problem of human face recognition, but the image database is restricted to human frontal faces with only slight illumination and expression changes. This paper applies the sparse representation based algorithm to the problem of generic image classification, with a certain degree of intra-class variations and background clutter. Experiments are conducted with the sparse representation based algorithm and Support Vector Machine (SVM) classifiers on 25 object categories selected from the Caltech101 dataset. Experimental results show that without the time-consuming parameter optimization, the sparse representation based algorithm achieves comparable performance with SVM. The experiments also demonstrate that the algorithm is robust to a certain degree of background clutter and intra-class variations with the bag-of-visual-words representations. The sparse representation based algorithm can also be applied to generic image classification task when the appropriate image feature is used.

INTRODUCTION
The task of generic image classification involves two important issues. One is image representation, the other is classification algorithm. Many image representation methods have been developed using various global features. Region-based features have also been developed by segmenting an image into several locally uniform regions and extracting feature for each region. Recently, keypoint-based image features are getting more and more attention for computer vision area. Keypoints, also known as interest points or salient regions, refer to local image.
Sparse Based Image Classification With Bag-of-Visual-Words Representations

patches which contain rich information, have some kind of saliency and can be stably detected under a certain degree of variations. The extraction of keypoint-based image feature usually includes two steps. First, keypoint detectors are used to automatically find the keypoints. Second, keypoint descriptors are used to represent the keypoint features. Performance has been evaluated among several different keypoint detectors and descriptors (Mikolajczyk et al., 2005; Mikolajczyk & Schmid, 2005).

Corresponding to the different kinds of image representation methods, many classification algorithms were studied (Csurka et al., 2004; Fergus, Perona, & Zisserman, 2003; Jing et al., 2004; Li & Perona, 2005; Sivic & Zisserman, 2003; Zhang et al., 2007). Image classification models can be divided into two classes. One class is generative models. The representative work is constellation model (Fergus, Perona, & Zisserman, 2003) which is a probabilistic model for object categories. The basic idea of this model is that an object is composed of several parts that are selected from the detected keypoints, with the appearance of the parts, scale, shape and occlusion modeled by probability density functions. A Bayesian hierarchical model was proposed (Li & Perona, 2005) for natural scene categories recognition, which learns the distribution of the visual words in each category.

The other class of image classification models is discriminative models, which have been proved to be effective for object classification (Csurka et al., 2004). A support vector machine (SVM) with the orderless bag of keypoints image representation was demonstrated to be effective for classification of texture and object images (Zhang et al., 2007). Zhang et al. (2006) proposed a hybrid of nearest neighbor classifiers and support vector machines to achieve good performance with reasonable computational complexity. Lazebnik et al. (2006) presented a multi-layer bag of keypoints feature with modified pyramid match kernels, which demonstrated that a well-designed bag-of-features method can outperform more sophisticated approaches.

Recently, sparse coding has been used for the learning of visual vocabulary or codebook and image representation. Yang et al. (2009) adopted sparse coding instead of K-means cluster to quantize the image local features and proposed a linear spatial pyramid matching kernel using image representation based on sparse codes. Considering the mutual dependence among local features, Gao et al. (2010) proposed a Laplacian sparse coding method to learn the codebook and quantize local features more robustly.

In 2009, theory from sparse signal representation was applied to the problem of human faces recognition (Wright et al., 2009), in which a test face is represented as a sparse linear combination of the faces from the training set. A sparse representation based classification algorithm by computing $l_1$-minimization problem is proposed. The authors gave new insights into two important issues in face recognition: feature extraction and robustness to occlusion. Although good performance was obtained, the image database was strictly confined to human frontal faces with only slight illumination and expression changes. Detection, cropping and normalization of the faces were done beforehand.

Our work differs from Wright et al. (2009) mainly in the following three aspects.

1. **Application Domain**: We apply 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. No preprocessing needs to be done for each image.

2. **Image Feature Space**: Wright et al. (2009) studied on the low dimensional sub-space in the high dimensional pixel space, which required image pixels to be strictly aligned. The paper argued that even downsampled images and random projections can do as
Related Content

An Interactive Visualization of Genetic Algorithm on 2-D Graph
www.igi-global.com/article/interactive-visualization-genetic-algorithm-graph/67997?camid=4v1a

A Formal Knowledge Retrieval System for Cognitive Computers and Cognitive Robotics
www.igi-global.com/article/a-formal-knowledge-retrieval-system-for-cognitive-computers-and-cognitive-robotics/101317?camid=4v1a

Machine Learning for Automated Polyp Detection in Computed Tomography Colonography
www.igi-global.com/chapter/machine-learning-automated-polyp-detection/56177?camid=4v1a

Empirical Studies on the Functional Complexity of Software in Large-Scale Software Systems
www.igi-global.com/article/empirical-studies-functional-complexity-software/60747?camid=4v1a