Encoding Optimization Using Nearest Neighbor Descriptor

Muhammad Rauf, Yongzhen Huang, and Liang Wang

National Lab of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract. The Bag-of-words framework is probably one of the best models used in image classification. In this model, coding plays a very important role in the classification process. There are many coding methods that have been proposed to encode images in different ways. The relationship between different codewords is studied, but the relationship among descriptors is not fully discovered. In this work, we aim to draw a relationship between descriptors, and propose a new method that can be used with other coding methods to improve the performance. The basic idea behind this is encoding the descriptor not only with its nearest codewords but also with the codewords of its nearest neighboring descriptors. Experiments on several benchmark datasets show that even using this simple relationship between the descriptors helps to improve coding methods.

Keywords: Nearest neighbor descriptor, Group saliency coding, Soft coding, Local constraint linear coding.

1 Introduction

One of the most important research areas in computer vision is image classification. There are different kinds of techniques used to serve this purpose. All these techniques have their benefits and drawbacks. Some work well on one kind of dataset and others can perform better on other kind of datasets. In all these techniques, the most commonly used framework is the Bag-of-words framework (BoW)[1][2]. This model consists of several steps, which starts from feature extraction and ends with classification. The hierarchy of these steps is, after feature extraction a codebook is generated and followed by feature coding, and before the classification feature pooling is performed.

All these steps have their own importance in the whole process of image classification using BoW. In recent years, encoding attracts lots of attention. There are different kinds of encoding methods that have been introduced to get better performance. Recent work[4][5] show that different coding methods perform different, even under the same framework. Soft voting outperforms hard voting[1] and the fisher kernel[6] has better performance than soft voting [3] with the same number of code words. These three are voting based methods and if we compare these voting based methods with reconstruction based coding[4],

S. Li et al. (Eds.): CCPR 2014, Part II, CCIS 484, pp. 81–88, 2014.

like local constraint linear coding (LLC)[7], we find that LLC has better results than the voting based coding methods. On the other hand the saliency[8] and group saliency coding[9] methods have implementation advantages over the reconstruction based coding, and perform faster than LLC. There are other coding methods introduce to improve the performance e.g., Laplacian sparse coding[10], multi-layer group sparse coding[11], improved Fisher kernel coding[12], Local tangent-based coding methods[13] and many more.

One thing that is common in all these methods is to encode one descriptor with codewords. In this process, we exactly do not know the relationship between a descriptor and its adjacent descriptors. If the descriptor extraction is not very dense then what are the influence of one descriptor to its neighboring descriptors and their codewords, i.e., the codewords used to encode descriptors. The main focus of our work is, to encode the descriptor by using the nearest neighbor descriptor's (NND) codewords and observe the change in performance. We explore a relationship between descriptors and by using this relationship, we update the codewords of descriptors. Our proposed technique is very simple and easy to implement.

The rest of the paper is arranged as follows. In Section 2 we introduce our proposed method in detail. In Section 3 first we discuss the datasets and the coding methods, and afterwards we evaluate our proposed technique. At the end in Section 4 we present the conclusion and our future work.

2 Nearest Neighbor Descriptor

The proposed method not only considers the structure of K-nearest codewords to a descriptor, but also takes account of the structure of neighboring descriptor codewords. We present a new technique that uses the descriptor-to-descriptor relationship during the encoding process. Results show that the locality of the descriptors has a very important role in encoding.

Our implementation is done in two different phases. First, we find K-nearest codewords of a descriptor and finally we update each descriptor's codewords based on the NND codewords. Let $X = [x_1, x_2, ..., x_N] \in R^D \times N$ be N D-dimensional descriptor form an image, and $B = [b_1, b_2, ..., b_M] \in R^D \times M$ be a codebook with M codewords.

2.1 Local Code Assignment

In this phase, we encode the descriptor with K codewords using the existing encoding methods. K is set to be a small number[20] and $[b_1, b_2, ..., b_K]$ is K closest codewords of x e.g., K=3 in Fig. 1(a). This is the local assignment of the nearest codewords to the descriptor. In the next phase, we generate new codewords that is based on the descriptor's and its neighboring descriptor's codewords.