Supervised incremental feature coding for SAR image classification

Shiyong Cui, Mihai Datcu
Remote Sensing Technology Institute (IMF)
German Aerospace Centre (DLR)
Emaill: shiyong.cui, mihai.datcu@dlr.de

Introduction

Recent years have witnessed a fruitful development of image representation. The most prevalent one is the Bag-of-Words (BoW) method, which gives state-of-the-art performance in many applications. It has four steps: local feature extraction, dictionary learning, feature coding, and feature pooling. In this paper, we focus on feature coding. On the basis of an analysis of currently popular feature coding methods, we propose a supervised incremental coding method. The most different characteristic of this method is that coding of a new image relies on the coding of the previous image from the same class. Therefore, we need to know the label of one image before coding. This point can be argued as a drawback of this method. However, we demonstrate that it could give much better feature for image classification. This finding gives some hints about further development of feature coding method. We believe that the entire class should be considered when coding the local features.

BoW feature extraction

BoW model has shown its significantly powerful ability in image classification. The framework of BoW model shown in Fig.1 has mainly four components: local feature extraction, codebook learning, feature coding and feature pooling. In this paper, we focus on feature coding. Based on the analysis of current popular feature coding methods, we propose a supervised incremental coding method.

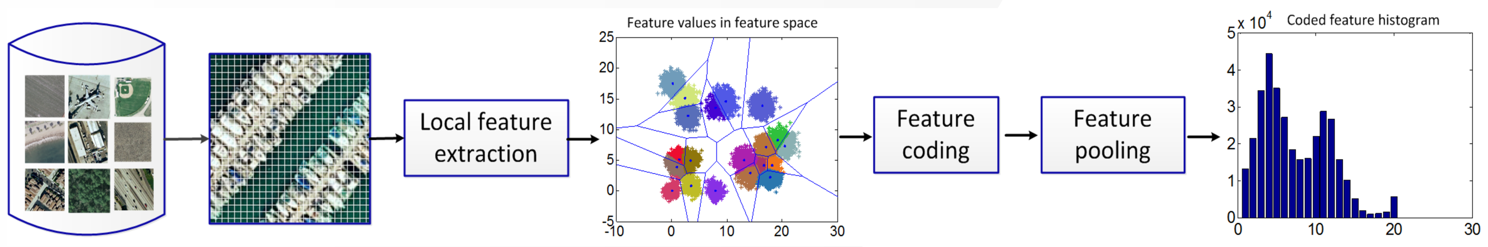


Figure 1. The framework of Bag-of-Words feature extraction.

Sparse feature coding

Sparse feature coding has been developed, as a generalization of *K*-means clustering , for image classification. Given a set of local descriptor x_i, It is formulated as

$$\min_{\mathbf{D},\mathbf{C}} \sum_{m=1}^{M} ||\mathbf{x}_m - \mathbf{D}\mathbf{c}_m||^2 + \lambda||\mathbf{c}_m||_1 \quad \text{subject to} \quad \|\mathbf{d}_k\|_2 \leqslant 1 \quad \forall m = 1, 2, ..., k$$

Where $\mathbf{D} = (\mathbf{d}_1, ..., \mathbf{d}_K) \in \mathbb{R}^{D\times K}$ is the codebook, and $\mathbf{c}_m \in \mathbb{R}^K$ is the reconstruction coefficient of the x_m. Compared with *K* – means, the constrained term has been relaxed from L0 norm to L1 norm. In the learning phase, both the codebook and the reconstruct coefficients are learned. In the coding phase, each local descriptor is coded using the learned dictionary. The word frequency histogram is replaced by a max pooling within each region in the framework of SPM.

Locality constraint feature coding

It is observed that non-zero coefficients are often assigned to the nearby elements in the codebook, so local version of the sparse coding (LLC) has been proposed. Different from kernel codebook, it is assumed that the current feature point for coding can be reconstructed using the *K* nearest clusters. The reconstruct coefficients can be computed by solving a least square problem. The weights for the rest clusters are set to zeros. It is formulated as follows

$$\min_{\mathbf{D},\mathbf{C}} \sum_{m=1}^{M} ||\mathbf{x}_m - \mathbf{D}\mathbf{c}_m||^2 + \lambda||\mathbf{s}_m \odot \mathbf{c}_m||_1 \quad \text{subject to} \quad \mathbf{1}^T\mathbf{c}_m = 1 \quad \forall m = 1, 2, ..., k$$

Where $\odot$ is the element wise product, and s_m is a vector of distances between the descriptor x_m and all elements in the dictionary. Different from sparse coding, LLC has some attractive properties, like better reconstruction, local smooth sparsity, etc. One of the most important one is that there is an analytical solution given by

$$\hat{\mathbf{c}}_m = \Big((\mathbf{D} - \mathbf{1}\mathbf{x}_i^T)(\mathbf{D} - \mathbf{1}\mathbf{x}_i^T)^T + \lambda \mathrm{diag}(\mathbf{d})\Big),$$
$$\mathbf{c}_m = \hat{\mathbf{c}}_m / \mathrm{sum}(\hat{\mathbf{c}}_m)$$

Supervised incremental feature coding

All the feature coding methods that were reviewed previously encode the local descriptors independently, which means that the images seen before have been forgotten. This is obviously in contradiction to the human learning. ***For instance, when you see a dog first time, you are told that that is a dog and you can encode the dog using simple features. After seeing dogs many times, obviously you should encode the dog better than the first time on the basis of previous knowledge learned about the dog***. By analogy, the feature coding algorithm should encode the local features in an incremental manner.

Based on this idea, we propose an incremental feature coding based on the LLC feature coding, as shown in Fig. 2 The main difference from previous encoding methods is that the encoding of the image depends on the previous encoded images and that the encoding is performed in an incremental manner. The algorithm starts from the database and a dictionary is learned using *k*-means clustering. The feature encoding is done class by class. The encoding starts from the first image in one class and the *k* nearest neighbors are searched. After that, the local reconstruction coefficients are obtained by solving the least square problem. Based on the reconstruction coefficients of the previous image, the encoding of the next image is performed by replacing only few coefficients in each column in **C**. Therefore, the reconstruction coefficient is learned in an incremental manner. For each image, the final feature vector of one image is extracted by max pooling.

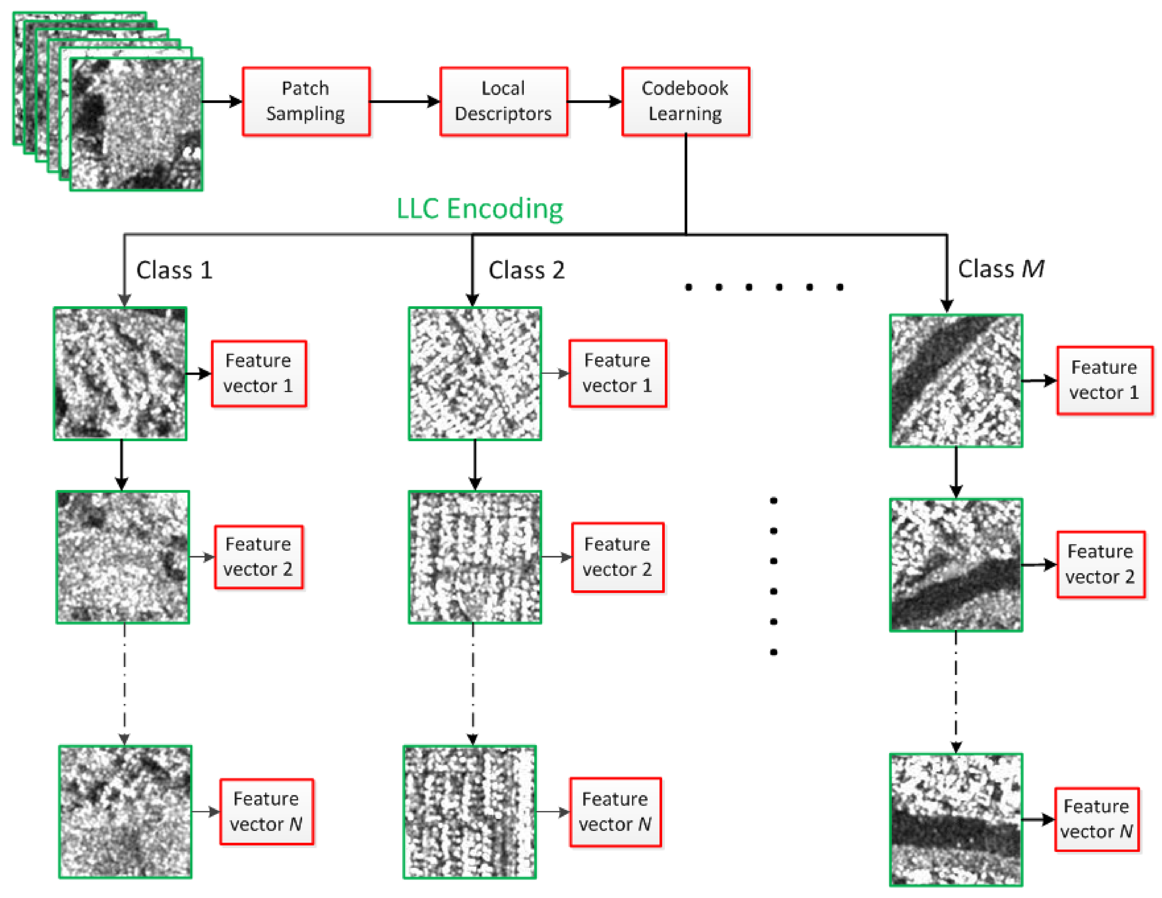
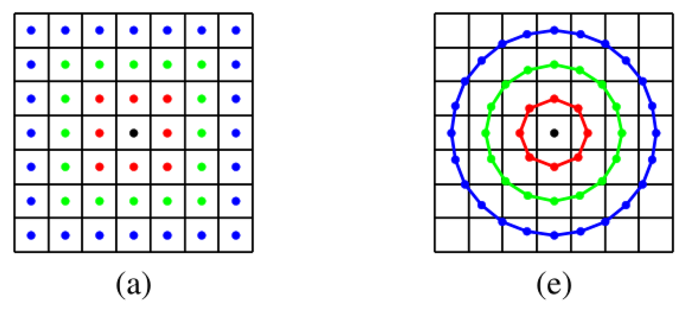


Figure 2. Supervised incremental feature coding.

Evaluation

- Dataset: 15 classes, 3434 SAR images
- Local features: SRP Global and SRP Angular-diff
- Comparison with vector quantization, kernel codebook coding, fisher vector, and LLC.



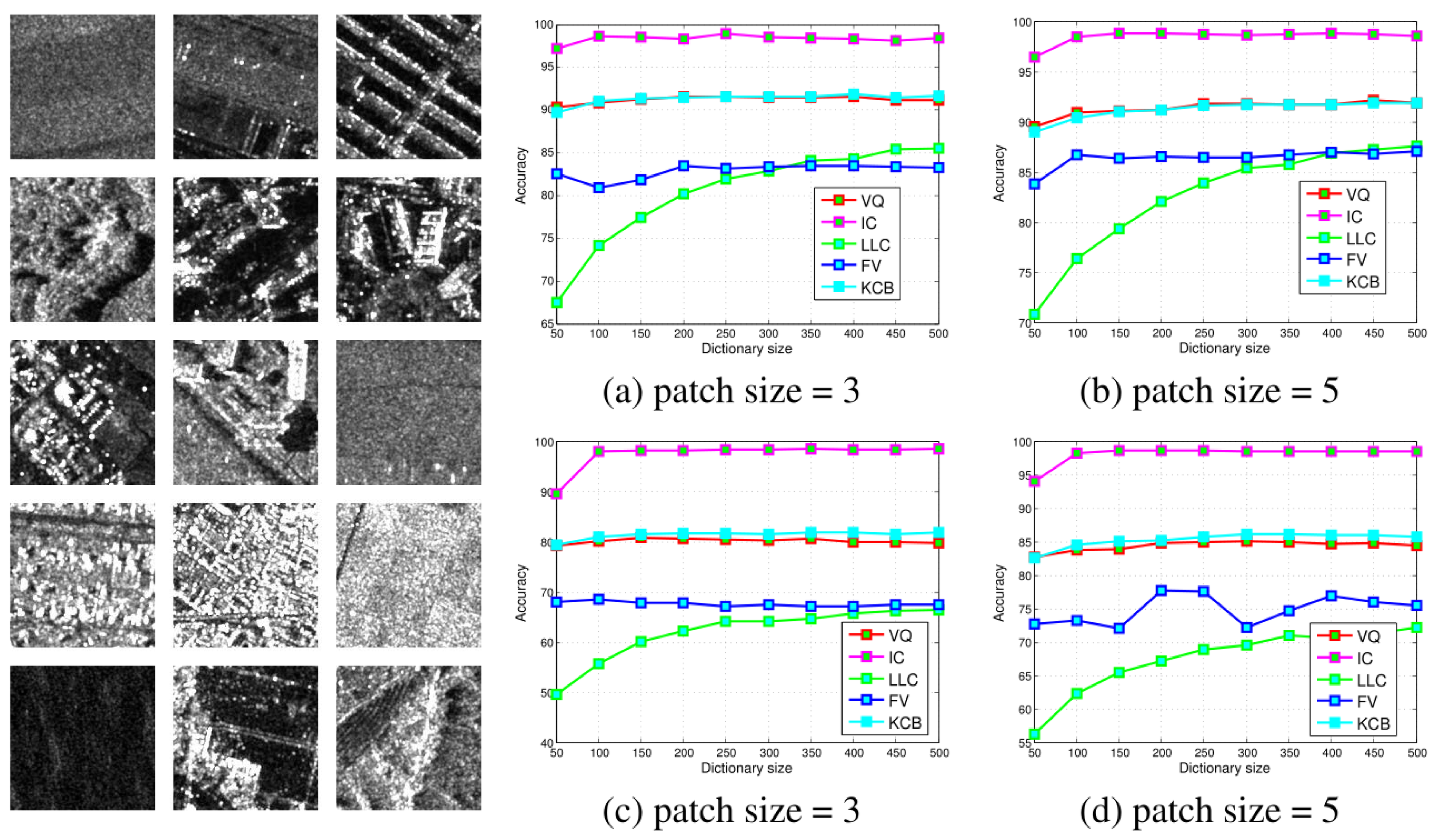


Figure 2. Test dataset with 15 classes and results.

Conclusion

In this paper, we propose a supervised incremental coding method. We argue that the coding method should encode the local feature in an incremental manner. The most different characteristic of this method is that the coding of a new image relies on the coding of the previous image from the same class. Therefore, we need to know the label of one image before coding. This point can be viewed as a drawback of this method. However, we demonstrate that it could give much better feature for image classification. This finding gives some hints about further development of the feature coding method. We believe that the entire class should be considered when coding the local features. In addition, this method can be applied in order to enhance the feature space when we have weak labels, like cascade learning.

- J. Yang, K. Yu, Y. Gong, and T. S. Huang, "Linear spatial pyramid matching using sparse coding for image classification," CVPR 2009.
- J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained Linear Coding for Image Classification," CVPR 2010
- L. Liu, P. Fieguth, D. Clausi, and G. Kuang, "Sorted random projections for robust rotation-invariant texture classification," Pattern Recognition, vol. 45, no.6, pp.2405-2418, 2012

REMOTE SENSING TECHNOLOGY INSTITUTE