Experimental Evaluation of the Bag-of-Features Model for Unsupervised Learning of Images

Mariana Afonso
marianafza@fe.up.pt
Department of Electrical Engineering,
University of Porto

Luis F. Teixeira
luisft@fe.up.pt
INESC TEC and Department of Computer Science,
University of Porto

The Bag-of-Features (BoF) [2] is a popular model that aims to represent images as an orderless collection of features without the use of any spatial information. Each image is represented by a frequency histogram of visual words from a codebook. Although the model is quite simple with regards to the implementation, there are several steps in which parameters and algorithms need to be chosen.

This work aimed to assess the performance of this model for the application of unsupervised learning for a set of images, also called image clustering. Additionally, it aims to provide valuable insight on the different steps of the model and to compare different algorithms in order to achieve the best performance for a given dataset. All the source code of this work is available open-source [1] at Github.

The applications of image clustering are endless and could include social network mining, more specifically for summarization of the huge amount of content shared everyday by millions of users.

In order to obtain the BoF representation of an image collection, many steps are required and are illustrated in Figure 1. The first one is the image description step, in which the input images are processed by first detecting keypoints or patches and then describing them using a certain algorithm. The next step is codebook learning, where a portion of the keypoints extracted from the images are clustered in order to obtain a codebook of visual words. This usually requires sampling of the total number of keypoints obtained from the images. The following step is the BoF representation of the images where each image is represented by a histogram of frequencies of visual words from the codebook obtained previously. The words are then filtered and the histograms are normalized following a chosen methodology. Finally, clustering is applied to the final representation of the images.

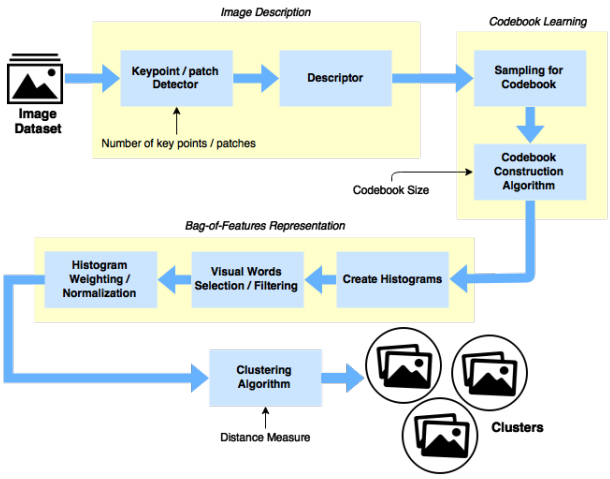

Figure 1: The process and the main steps of the Bag-of-Features model for the application of image clustering.

Due to the popularity of the BoF model, a number of works have been focused on evaluating its performance. For instance, in [1] the authors presented the results of an experimental study concerning the BoF model applied to the problem of image classification. Several key steps of the model were tested using different algorithms and parameters. Another empirical study presented in [3] evaluated the impact of applying techniques used in text categorization to the BoF model for the application of scene classification.

The main contributions of this study are: (1) the experimental analysis of the BoF model for image clustering, (2) the addition of a number of steps and algorithms (e.g. sampling the features for codebook learning and visual words selection) and (3) the proposal of a sampling technique for the features obtained from the images for the codebook learning procedure.

In terms of the experimental design, in each of the steps of the model, several parameters and algorithms were varied. Three datasets were used in order to obtain different levels of difficulty and complexity, an object dataset and two scene datasets. The performance measures used were the Normalized Mutual Information (NMI) and the Adjusted Rand Index (ARI). Both these metrics are popular choices for assessing clustering results and since they have different natures, a more complete evaluation was possible. All the tests were repeated 10 times to obtain the average and the standard deviation of the performance measures.

The results from the detectors and descriptors step indicate that the performance of the BoF model highly depends on the algorithm for the description of the images and is less influenced by the detector used (except for object datasets). Also, the choice of the best algorithms is dependent on the dataset.

Next, the influence of the average number of keypoints per image and the codebook size were tested varying those values for the three datasets. The results show that regardless of the codebook size used, the performance almost always increases with the average number of keypoints per image. Additionally, it could be observed that the ideal size of the codebook increases with the complexity of the dataset.

It was also found that the strategy for sampling the keypoints, the codebook learning algorithm, the visual words filtering and the normalization and weighting of the histograms were steps that did not influence the performance of the model significantly, and therefore, do not require much tunning of the parameters and algorithms used.

Lastly, different clustering algorithms were tested for the final clustering of the histograms and the conclusion of this last step is that, although an algorithm which does not require the number of clusters is desirable, they perform much worse and require several parameters that need to be adjusted, which can be very specific to a given set of images and settings.

The final performance of the image clustering task for the three datasets using the BoF model after the tunning of the parameters and algorithms can be found in Table 1.

Table 1: Final performance of the BoF model for the three datasets tested.

Dataset	Avg. ARI	Avg. NMI
Coil-20 (object - low difficulty)	**67,4%**	**84,8%**
Natural and Urban (scene - medium difficulty)	**30,2%**	**40,6%**
Event (scene - high difficulty)	**19,4%**	**27,4%**

From all the different experiences developed and presented in this work, it can be concluded that although the Bag-of-Features model can be successfully applied to the problem of unsupervised learning images, it provides a poor representation of the images when the datasets represent complex scenes and requires fine tunning of the algorithms and parameters used in each step. For this reason, more advanced techniques are required in order to be able to effectively extract information from large image collections in an unsupervised fashion.

[1] Eric Nowak, Frédéric Jurie, and Bill Triggs. Sampling strategies for bag-of-features image classification. In *Computer Vision–ECCV 2006*, pages 490–503. Springer, 2006.

[2] Josef Sivic and Andrew Zisserman. Video google: A text retrieval approach to object matching in videos. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1470–1477. IEEE, 2003.

[3] Jun Yang, Yu-Gang Jiang, Alexander G Hauptmann, and Chong-Wah Ngo. Evaluating bag-of-visual-words representations in scene classification. In *Proceedings of the international workshop on Workshop on multimedia information retrieval*, pages 197–206. ACM, 2007.

[1] https://github.com/marianaAfonso/BOFClustering