Fast Vocabulary Construction and Testing of Multi-class Texture Classification

Paheerathy[†]**, R., Ramanan**[†]**, A.** and **Niranjan**[‡]**, M.**

[†]Department of Computer Science, University of Jaffna, Sri Lanka.

[‡] School of Electronics and Computer Science, University of Southampton, UK.

r.pahee@gmail.com, a.ramanan@jfn.ac.lk, mn@ecs.soton.ac.uk

Introduction

The order less bag-of-keypoints or bag-of-features (BoF) approach has the advantage of simplicity, lack of global geometry, and state-of-the-art performance in visual object classification tasks. In such a model the construction of a visual vocabulary plays a crucial role that not only affects the classification performance but also the construction process is very time consuming which makes it hard to apply on large datasets.

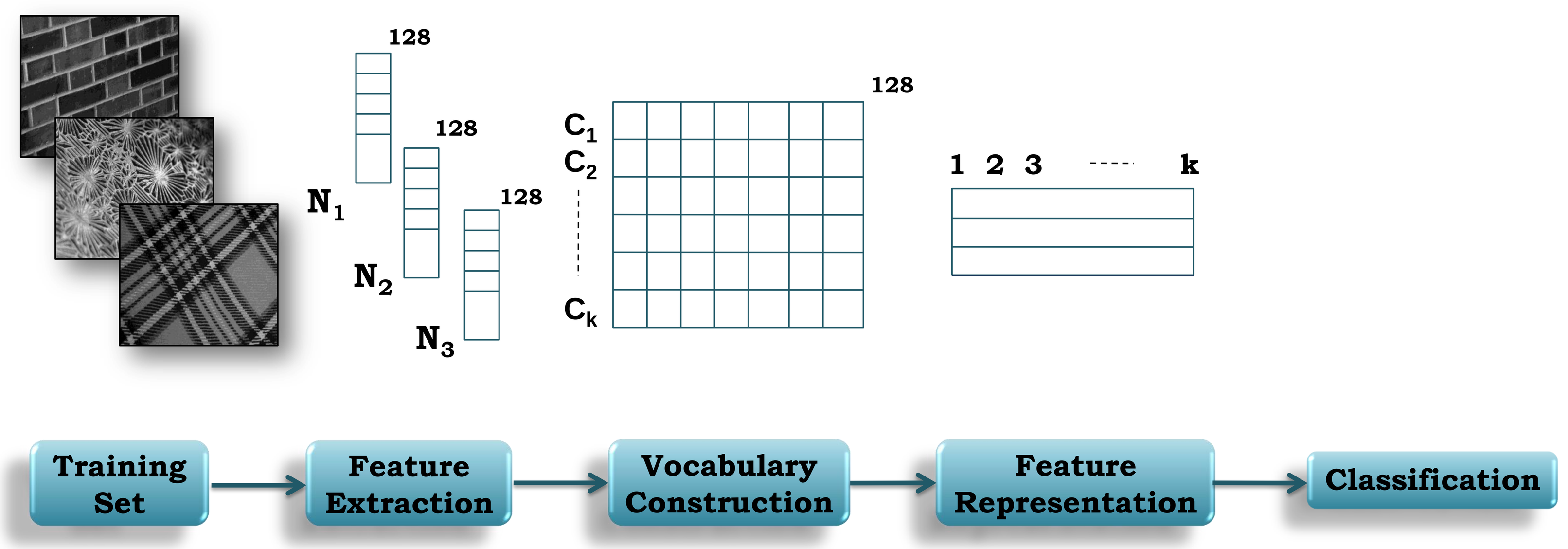


Figure 1: Major stages of a Bag-of-keypoints approach

Motivation

Our goal is to address two of the performance bottleneck in a BoF approach,

- The vocabulary construction in a bag-of-keypoints approach by means of a one-pass resource-allocating codebook (RAC) approach that drastically reduces the construction time while maintaining performance comparable to the state-of-the-art approaches.
- The multi-class classification by means of an unbalanced decision tree (UDT) which is based on a "knock-out" strategy with at most (N-1) classifiers to make a decision on any input pattern.

Methodology

Feature Extraction: We use the Scale-Invariant Feature Transform (SIFT) features which are computed as local histograms of edge directions computed over different parts of the interest region, giving a 128-dimension vector. This allows to capture the structure of the local image regions in more texture-like content.

Codebook Construction: The Resource-Allocating Codebook (RAC) technique was employed in constructing a codebook. RAC starts by arbitrarily assigning the first keypoint as an entry to the initial codebook. When a subsequent data item is processed, its minimum distance to all entries in the current codebook is computed, using an appropriate distance metric. If this distance is smaller than a predefined threshold, r, which is the radius of a hypersphere, the current codebook is retained and no action is taken with respect to the processed data item. If the threshold is exceeded by the smallest distance to centroids, a new entry in the codebook is created by including the current data item as the additional entry. This process is continued until all data items are seen only once.

Classification: We use the Unbalanced Decision Tree (UDT) SVMs for classification. Each decision node of UDT is an optimal classification model. The optimal model for each decision node is the one versus all (OVA) based classifier that yields the highest performance measure. Starting at the root node, one selected class is evaluated against the rest by the optimal model. Then the UDT proceeds to the next level by eliminating the selected class from the previous level. UDT terminates when it returns an output pattern at a level of the decision node. In contrast, we can say that UDT uses a "knock-out" strategy with at most (k-1) classifiers to make a decision on any input pattern.

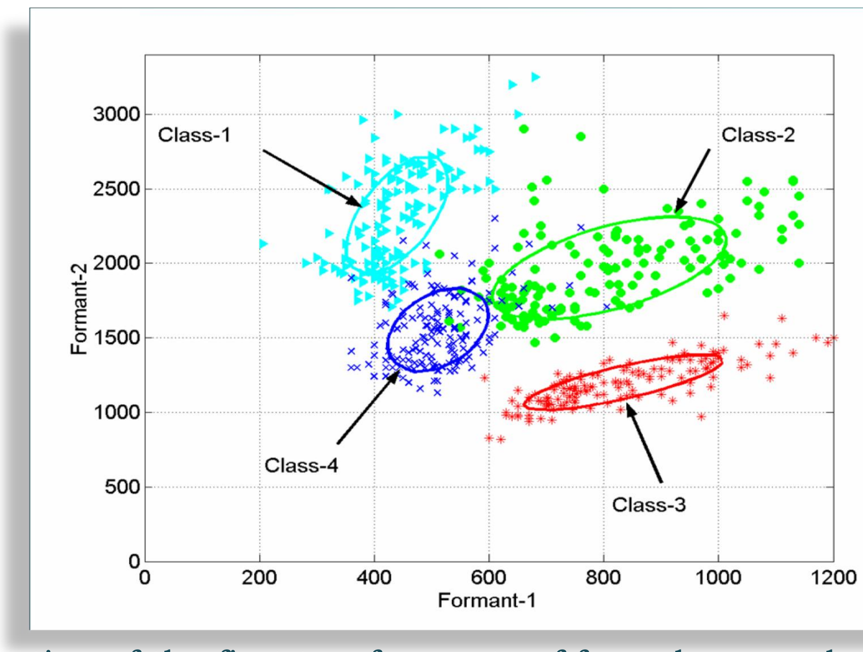


Figure 2: Distribution of the first two formants of four classes selected from the Peterson and Barney (1952)'s vowel dataset.

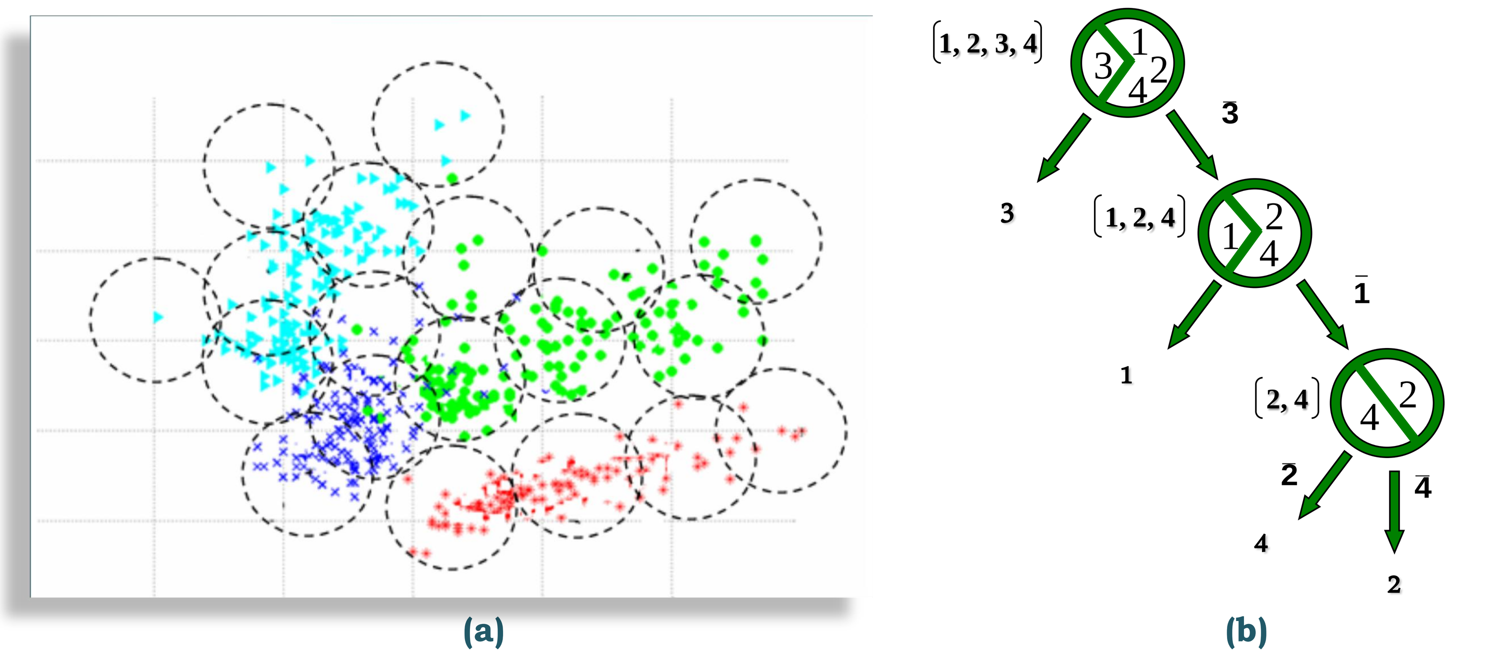
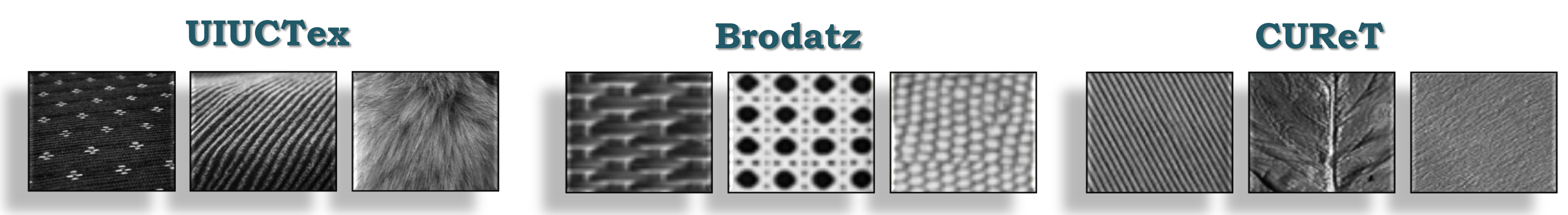


Figure 3: (a) Partitioning the feature space in Figure 2 using Resource-Allocating Codebook (RAC) approach and (b) corresponding UDT architecture and the classification problems at each node for finding the best class out of the four classes. The equivalent list state for each node is shown next to that node. The RAC has more unequal points that span more widely and capture rare points in the feature space.

Datasets and Experimental Setup

- The UIUCTex dataset contains 25 texture classes with 40 images per class for which a ten-fold cross-validation was used.
- The Brodatz dataset contains 111 texture classes with one image per class for which a three-fold cross-validation was used.
- The CUReT dataset contains 61 texture classes with 92 images for each class for which a two-fold cross-validation was used.



Results

Table 1: A comparison of time taken to construct a class-wise visual codebook for the UIUCTex dataset and the time taken for classifying unknown images using the UDT based SVM. The reported times (in seconds) are the average of the 10-fold cross-validation.

	Vocabulary Construction	Classification
K-means + DAG	4670 seconds	143 seconds
RAC + UDT	6 seconds	101 seconds
Ours(V500)	3 seconds	77 seconds

Table 2: A comparison of classification performance obtained by several methods applied on the UIUCTex, Brodatz, and CUReT datasets.

Method	UIUCTex	Brodatz	CUReT
Lazebnik et al. (2005)	96.4 ± 0.9	89.8 ± 1.0	72.5 ± 0.7
Nowak et al. (2006)	83.5 ± 0.8	91.0	-
Zhang et al. (2006)	98.3 ± 0.5	96.1 ± 0.8	95.3 ± 0.4
Soottitantawat et al. (2011)	93.6	90.8	-
Ours	96.9 ± 1.2	90.4 ± 1.4	72.2 ± 0.1

Discussion and Conclusion

- In the UIUCTex dataset, our approach performs better than the methods in [3], [4] and [8] but comparable results to [10].
- In the Brodatz dataset, our approach is comparable to the method in [3].
- The advantage of our method is that it achieves comparable performance to previously reported results in texture classification at a drastically reduced time.
- The reason behind the less performance on the CUReT dataset is due to the usage of patch-based descriptors. Since most of the CUReT textures are very homogeneous and high-frequency, lacking salient structures such as blobs and corners, keypoint extraction does not produce very good image representations.

Reference

[1]. P. Brodatz, Textures: A Photographic Album for Artists and Designers, Dover, New York, 1966.
[2]. K. J. Dana, B. van Ginneken, S. Nayar, and J. Koenderink, Reflectance and Texture of Real-World Surfaces, in ACM Transactions on Graphics, volume 18, pp. 1–34, 1999.
[3]. S. Lazebnik, C. Schmid, and J. Ponce, A Sparse Texture Representation using Local Affine Regions, in IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), volume 27, pp. 1265–1278, 2005.
[4]. E. Nowak, F. Jurie, and B. Triggs, Sampling Strategies for Bag-of-keypoints Image Classification, in Proceedings of the European Conference on Computer Vision (ECCV), pp. 490–503, 2006.
[5]. G. Peterson and H.Barney, Control methods used in a study of the vowels, J. Acoustical Society of America, 24, pp. 175-184, 1952.
[6]. A. Ramanan and M. Niranjan. A One-pass Resource-Allocating Codebook for Patch-based Visual Object Recognition. In proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP), pp. 35-40, 2010.
[7]. A. Ramanan, S. Suppharangsan, and M. Niranjan. Unbalanced Decision Trees for Multi-Class Classification. In Proceedings of the Second IEEE International Conference on Industrial and Information Systems (ICIIS), pp. 291-294, 2007.
[8]. S. Soottitantawat and S. Auwatanamongkol, Texture Classification using an Invariant Texture Representation and a Tree Matching Kernel, In International Journal of Computer Science Issues (IJCSI), volume 8, 2011.
[9]. UIUCTex : http://www-cvr.ai.uiuc.edu/ponce grp.
[10] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study, In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), 2006.