Sequentially Constructing Discriminative Codebook with Classifier Training for Object Recognition

V. Vinoharan and A. Ramanan
Department of Computer Science, University of Jaffna, Sri Lanka.
{vinoharan, a.ramanan}@jfn.ac.lk

SLSWCS Student Workshop on Computer Science

Introduction

The bag-of-features approach is a popular technique for representing image content. In such a system a visual codebook plays a crucial role. Researchers cover a large-scale of training image set to construct a codebook. An important issue of the visual codebook representation is its discriminative power and dimensionality. This higher dimensionality curses the subsequent classifier training procedure.

In this work we investigate whether the use of increased number of training images will contribute significantly to improve the performance of classification or is it worth to focus on the selection of discriminative features and the development of better object models.

Objectives

- Does all training images contribute to the discriminative power of a codebook?
- To propose an incremental way of constructing a compact codebook while maintaining its discriminative power.

Methodology

- Figure 1 and Algorithm 1 describe the process of sequentially constructing a codebook using an extended resource allocation codebook (RAC) technique [1]. SIFT descriptors [2] were used in our experiment.
- The incremental approach of constructing a codebook is halted either all training images are processed or a desired classification rate is achieved.

Methodology ...

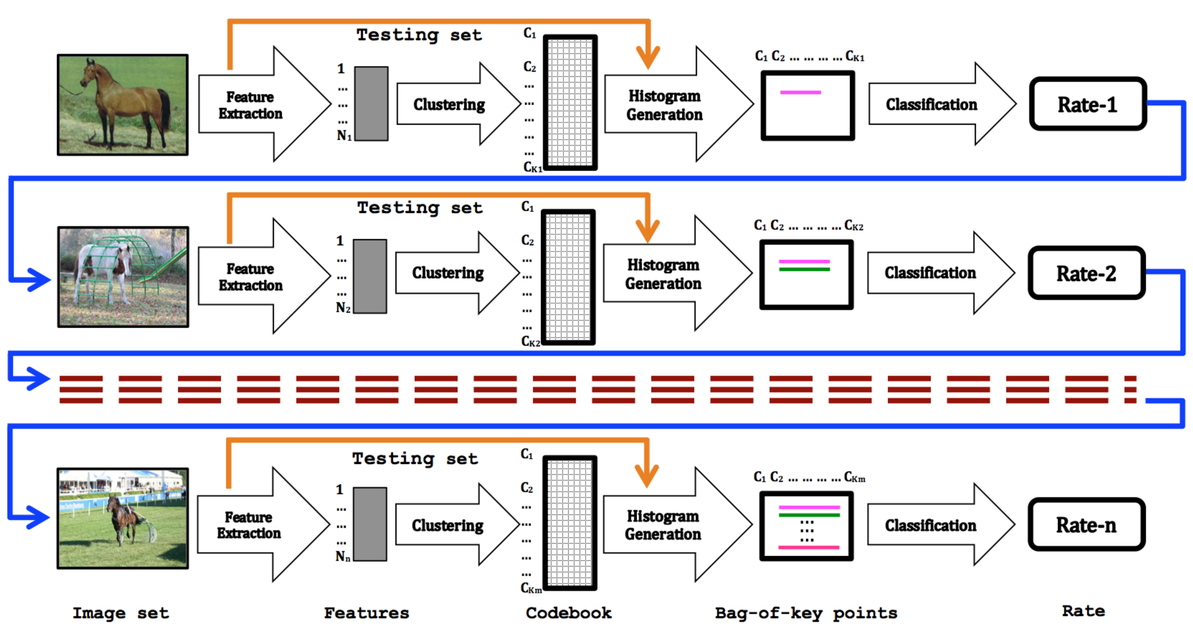

Figure 1: The overall framework of the proposed technique to sequentially constructing visual codebook for a object-specific category (e.g.horse)

Methodology ...

Algorithm 1: Sequentially constructing codebook image-by-image

Input: Training images (trImgs), Testing images (teImgs)
Output: Visual codebook (CB), Classification accuracy (rate)
Process:

```
for all img_i ∈ {trImgs, teImgs} do
    interestPts ← detectPts(image)
    descriptors ← describePts(interestPts)
end for
r ← predefined value
// Initialise the codebook CB
D ← descrips(img_1) //where img_1 ∈ trImgs
CB ← D_1
i ← 1
for all img_i ∈ trImgs do
    D ← descrips(img_i)
    j ← 1
    while (j ≤ size(D)) do
        if min‖ D_j − CB ‖^2 > r^2 then
            Create a new hypersphere of r such that,
            CB ← {CB ∪ D_j}
        end if
        j ← j + 1
    end while
    trainHist ← computeHist(CB, descrips(trImgs))
    testHist ← computeHist(CB, descrips(teImgs))
    rate_i ← classify(trainHist, testHist)
    i ← i + 1
end for
```

Experimental Setup

- We tested our approach on PASCAL VOC 2007 Challenge dataset [3]. It consists of 9963 images from 20 categories.
- SIFT features were clustered independently using K-means with K = 250 and extended RAC with r = 0.89.

Results

In our proposed technique:

- on average about 13-22% of the training images is only needed to construct a discriminative codebook.
- constructs a compact codebook which is around 60% size of the codebook constructed either by K-means method or RAC technique.

Results ...

Table 1: Classification rate with codebook size and number of training images for the standard RAC and proposed sequential learning method with r = 0.89.

Object	RAC			Ours		
	#imgs	CB	rate	#imgs	CB	rate
Aeroplane vs Bird	568	499	0.83	34	279	**0.87**
Aeroplane vs Boat	419	471	0.80	22	236	**0.80**
Aeroplane vs Horse	525	535	0.87	58	341	**0.90**
Aeroplane vs Sofa	467	473	0.87	90	356	**0.88**
Bicycle vs Motorbike	488	493	0.67	22	274	**0.68**
Bird vs Cat	667	449	0.73	72	315	**0.75**
Boat vs Bus	367	429	0.82	46	303	**0.84**
Boat vs TVmonitor	437	402	0.88	14	165	**0.89**
Bottle vs Pottedplant	489	380	0.64	110	348	**0.65**
Bus vs Train	447	464	0.70	86	392	**0.71**
Cat vs Dog	758	465	0.65	30	247	**0.65**
Chair vs Dog	866	484	0.81	32	255	**0.81**
Cow vs Sheep	237	300	0.63	42	233	**0.65**
Diningtable vs Pottedplant	445	384	0.61	34	263	**0.64**
Pottedplant vs TVmonitor	501	398	0.68	132	373	**0.70**
Train vs TVmonitor	517	437	0.85	46	293	**0.86**

Discussion and Conclusion

- We optimise the process of constructing codebooks with less memory requirement and speeding-up the approach while maintaining compactness and discriminative power in recognition.
- Testing results shows that not all images are needed for constructing a discriminative codebook.
- Thus, this work suggests an alternative view to the research community working with the patch-based object recognition to emphasize retaining of more discriminative descriptors rather than the reminiscent of the 'BIG data' hypothesis.

References

[1] A. Ramanan and M. Niranjan. A one-pass resource-allocating codebook for patch-based visual object recognition. pages 35–40, 2010.

[2] D. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[3] M. Everingham *et al*. The pascal visual object classes challenge. *International journal of computer vision*, 88(2):303–338, 2010.