Improving Visual Codebooks Generation Using Additive Groves

Pablo D Robles-Granda
Computer Science Department, Purdue University

Abstract

The aim of this project is to analyze effectiveness and efficiency of visual codebook generation of a current image classification technique, Extremely Randomized Clustering Forests (ERCF), based on Leo Breiman's random forest. The objective is to use an adaptive feature extraction approach (Additive Groves) to capture more effectively the discriminative information of the images. The method shows how using sparse low level feature selection can provide better visual discriminative information than the iterative saliency formation of the original technique, and also a new tree-based organization of the saliency can offer improved accuracy in image retrieval.

Introduction

Image codebooks
Objective: Encode image's low level features with decision tree functions.

Image Sample ERF

$$F_t(\vec{x}) \to R^t,$$
$$\vec{x} \in R^s,\ s >> t$$

Background
Feature extraction is a critical step to reduce the dimension of the original images and provide effective representations for them. A recent approach to CBIR is to use extremely randomized clustering forests (ERCF) (Moosmann, F., et al. 2008) which is an extension of extremely random forests (ERF)

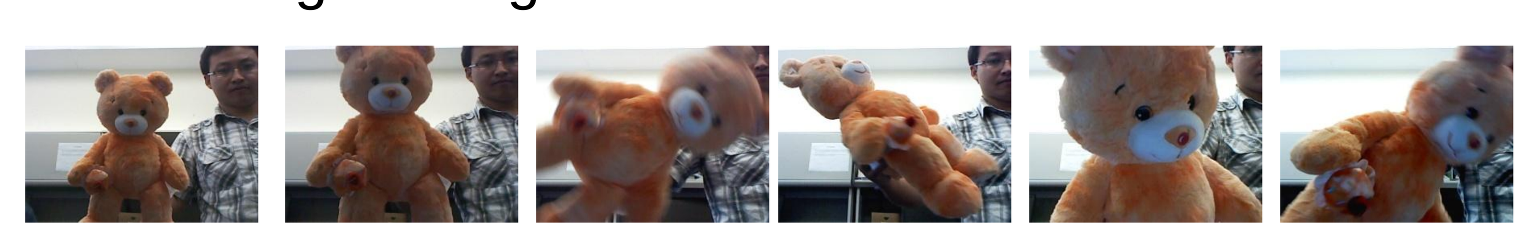
ERCF in a nutshell:
ERCF creates image codebooks using the procedure described in the figure above (which works as a feature space clustering method) and then a second machine learning (ML) classifier is applied for a final result:

Methods

For replication of experiments a version of ERCF for C++ MS Visual Studio 2005 was developed. Additional modules were incorporated in order to test different hypothesis. The datasets used for experiments have been divided in two categories: private datasets and public datasets. Depicted below, the private datasets were used to test classification speed against accuracy. Public databases, including Graz -02 and Caltech 256, have been tested partially. The experiments with the private dataset were performed to measure the both speed and accuracy characteristics of the method with various types of images. The ERCF method was compared to ERF alone. For that reason, two private datasets were generated.

Private Dataset 1: 12 categories, and 41 features per image. 1 image per category was used during training

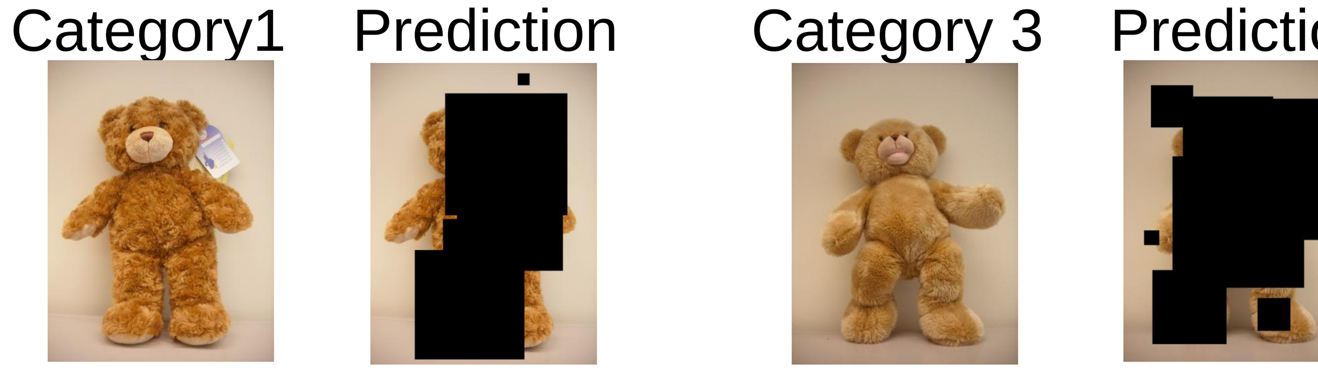
Private dataset 2: 5 categories and 9 features per image (used for video analysis). 9 images per category were used during training



Preliminary results

Quantitative results.
The first private dataset was created with 12 categories of objects and 41 features per image. The system was later tested with a test dataset containing various objects belonging to each of the 12 original categories The performance was very high with an accuracy of >90% for 80% of images in an average time of 0.87 seconds.
The second dataset was created with 5 different categories of objects and 9 features per image. This dataset was tested with ERF to categorize images from video source containing objects belonging or not to the 5 categories. The classification task took less than a second in average to categorize the entire scene (total of 60 images in one minute).

Category1 Prediction Category 3 Prediction



Error rate:

45 images from video source were tested using additive ensembles of trees. The classification accuracy is described in the confusion matrices.

ROC features

	1	2	3	4	5
1	79	49	4	3	
2	2	121	9	3	
3		5	107	23	
4		3	23	109	
5		8	21	27	79

Error rate
26.67%

ROC images

	1	2	3	4	5
1	7	2			
2		9			
3			7	2	
4			2	7	
5			2		7

Error rate
17.78%

References

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32.
Mossmann, F., Nowak, E. & Jurie, F. (2008). Randomized Clustering Forests for Image Classification. In IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 30, NO. 9, (pp. 1632-1646).