

# B657 Computer Vision

Interim Progress Report

## Final Project

Srikanth Srinivas Holavanahal

Pradeep Kumar Ravilla

Manikandan Murugesan

Spring 2016

# Introduction

## **Scene classification using a multi-resolution bag-of-features model**

Our initial plan was to implement the interesting problem of scene classification and automatically labeling the image among a set of semantic categories. There exist lots of scope for this project across various domains. For instance, it can be used to check whether the pictures uploaded to Yelp/Foursquare/Trip Advisor were sorted in the correct category.

To implement this project we referred multiple papers and combined multiple approaches to come up with a scene classifier which classifies with a near perfect accuracy. We started off trying to implement a multi-resolution bag of features model. This we achieve by resizing the images into a set of 3 different resolutions and then computing a set of features and converting them into a visual codebook by implementing k-means clustering. We hoped to train this data using a convolutional neural network and then classifying the given query image using this trained data.

## Background and related work

We had referred a paper titled “Scene classification using a multi-resolution bag-of-features model”<sup>[1]</sup> which discusses the procedure we decided to implement. The paper discusses in detail the steps to do multi resolution bag of features model. To implement k-means clustering, we referred another paper titled “An efficient k-means clustering algorithm: analysis and implementation”<sup>[2]</sup>. We got our dataset from the resources mentioned in the paper “Scene classification using a multi-resolution bag-of-features model”. Since part of this work had been done for Assignment 3, we have decided to revise our plan and come up with different work.

# Revised research plan

Our initial plan was to implement the bag of features model and then to train our CNN with the visual codebook got using the bag of features model. When we researched the feasibility of this approach, we observed that it is not possible to train our CNN on these features. So we have decided to stick to implementing the scene classification approach discussed in the paper and then to analyze in depth why this approach works so well for scene classification. Also we are trying to come up with certain approaches which might give us better results.

## Progress so far

We have implemented a naive version of the paper and we are getting a decent accuracy. We have to incorporate the minor details that we have missed while implementing the paper and we have to play around with the magic parameters to get better accuracy. Once we do this, we have to do a qualitative analysis on how well our approach works and what kind of images our approach works well for and across the kind of images for which our approach fails.

We ran our naive implementation across a small dataset of 6 classes containing 100 test images and then we classified the test set of 150 images into these 6 classes. The results we have obtained so far can be found in the table that follows. This is just a sample dataset and the dataset that we are going to implement for the final system is going to be larger than this.

	Casino	Child Room	Church	Classroom	Cloister	Closet	
Casino	5	0	19	1	0	0	
Child Room	0	6	16	0	3	0	
Church	0	3	19	3	0	0	
Classroom	2	8	1	14	0	0	
Cloister	3	3	0	0	19	0	
Closet	4	2	4	11	0	4	

The current accuracy we are getting is 44%

## REFERENCES:

- 1 - "Scene classification using a multi-resolution bag-of-features model" by Li Zhou, Zongtan Zhou, Dewen Hu  
<http://www.sciencedirect.com/science/article/pii/S0031320312003330>
- 2 - "An efficient k-means clustering algorithm: analysis and implementation" by D. M. Mount ; N. S. Netanyahu ; C. D. Piatko ; R. Silverman ; A. Y. Wu  
<http://ieeexplore.ieee.org/xpl/abstractAuthors.jsp?arnumber=1017616>
- 2 - "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006" by S. Lazebnik, C. Schmid, J. Ponce  
[http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=1641019](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1641019)
- 3 - "Scene categorization via contextual visual words" by Jianzhao Qin, Nelson H.C. Yung  
<http://www.sciencedirect.com/science/article/pii/S0031320309004245>