Object Recognition

Group no. L8
Gaurav Arya
Rajesh Dhania
Mentor : Dr. Amitabha Mukherjee
Dept. of Computer Science and Engineering
{aryag,rjdhania,amit} @ iitk.ac.in
April 13-04-2012

Abstract:

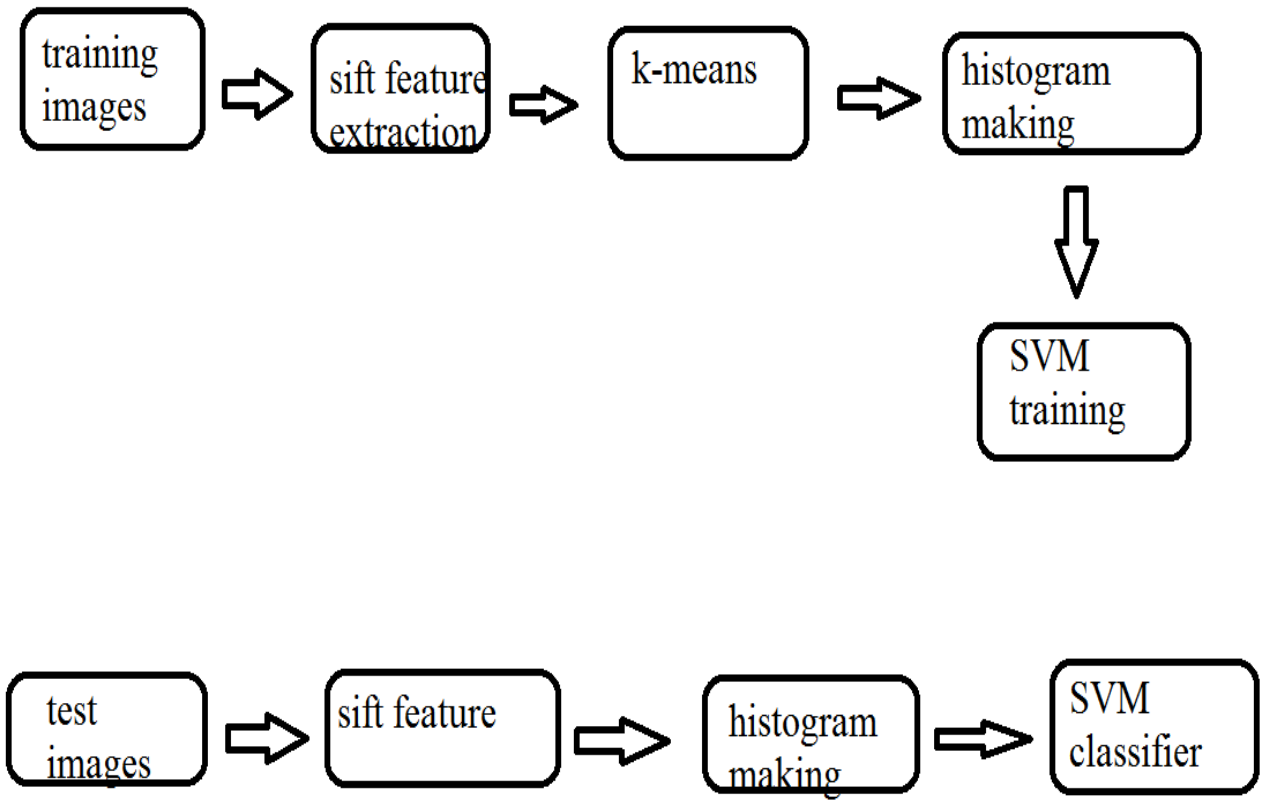
An object recognition system finds objects in the real world from an image of the world, using object models which are known a priori. .

The object recognition problem can be defined as a labeling problem based on models of known objects. Formally, given an image containing one or more objects of interest (and background) and a set of labels corresponding to a set of models known to the system, the system should assign correct labels to regions, or a set of regions, in the image.

Our Approach:
we are following bag of words approach to achieve our goal.

Implementation Details:
The entire implementation is in MATLAB. Firstly both the model and the scene images are converted to gray scale(.pgm (required for sift descriptor)).Thanks to the IrfanViewer for their image convertor software .*www.irfanview.com/*

Flow Diagram :



Shape Invariant Feature Transform (Sift) Feature:
We are extracting SIFT of every image .Now each image is represented as a n X 128 dimension feature vector.

K means:
when we extract all of the feature vectors of each images we concatenate them all vertically and apply k means algorithm on the total feature vector and compute the centroids (k codewords) of the clusters.we are taking here the k=128 .

Build Histograms of Training Images
We converted each training image into a histogram that counts the number of times each visual vocabulary word was found in the image. For every SIFT feature found in an image, we found which of the 128 cluster centers it was closest to, and added 1 to that word's histogram count. We then normalized each histogram, so that the sum of all the bins' entries is 1.Thanks to R. Bunschoten for their euiclidean distance code

SVM generation and classification:
We fed all of the train images' histograms (labeled with the scene class of the image) into a linear SVM .We are then classifying our test image using one vs. one SVM model. We got a total of ${}^{n}C_{2}$ SVM classifiers, now we are counting the frequency of the output class and we assign test image to that classifier which have highest frequency.

Results:
We took 40 images(.jpg) of each of the 5 classes from the James Z.Wang Research Group database (http://wang.ist.psu.edu/.../test1.tar) (mountain,dinosaur,flower,dishes,bus) for the training and 7 images each (from google images) for testing .we tested it for 3 different values of k.

We are getting accuracy near about 65% for k= 128. (excluding the ties)
mountain = 5/7
dinosaur = 5/7
bus = 5/7
flower = 4/7
dishes = 4/7

for k = 64 nearly 60%
mountain = 5/7
dinosaur = 4/7
dishes = 4/7
flower = 4/7
bus = 4/7

for k = 32 nearly 60%
mountain = 4/7
dinosaur = 5/7

dishes = 4/7
flower = 4/7
bus = 4/7

References:

1.)Sift Feature Extraction by David

2).Li Fei-Fei (UIUC) .Bag of words model,*people.csail.mit.edu/torralba/shortCourseRLOC/slides/part_1.**ppt***

3.) WikipEdia