

Vision Based Outdoor Localization of IIITD Campus

• • •

CV Final Presentation

Srinidhi Hegde 2013164

Tarun Malhotra 2013112

Neeraj Vaishnao 2013142

Introduction

- **WHAT ?**
 - To estimate the GPS location (outdoors) of the user in the IIIT-Delhi Campus from images
- **WHY ?**
 - Applications - **Egocentric Localization, Oil Pipeline Inspection, Mine Exploration, Military Applications, Crime Scene investigation**
 - Motivated from “**Where am I ?**” by ICCV, Computer Vision Contest
- **HOW ?**
 - That’s what we will see in this presentation.

Location Recognition - Feature Extraction

- SIFT Descriptors - Basic Idea : invariance to geometric transformation and illumination
 - extracts blob like feature points and describe them with a scale, illumination, and rotational invariant descriptor.
 - does not give an overall impression of the image (Not a global descriptor).
 - But, for recognition, Global descriptor is needed. Solution : Bag Of Features
- Bag of Words descriptors -
 - create a vocabulary of features with k words
 - this partitions the continuous SIFT feature space into k regions
 - represent images as bags of quantized SIFT features, based on the vocabulary

Motion Estimation

- Pick 2 best candidate images based on number of matches.
- GPS estimation based on number of matches among inliers as:

$$\frac{N_{ref1}P_{ref1} + N_{ref2}P_{ref2}}{N_{ref1} + N_{ref2}},$$

- If only one image in inliers then assign the GPS of the best match image.
- More accurate GPS estimation using structure of motion techniques of triangulation - uses epipolar constraints.

Vocabulary

Formed the vocabulary by sampling many local features from our training set and then clustering them using k-means.

- The number of k-means clusters is the size of our vocabulary and hence the size of our new feature space.
 - Tried for 100,200 and 1500 bags
- Clustering is a time consuming process. Built the vocabulary once, and stored the centroids of the clusters.
- For any new SIFT feature we observe, we can figure out which region it belongs to using the saved centroids of our original clusters.

Training Classifiers

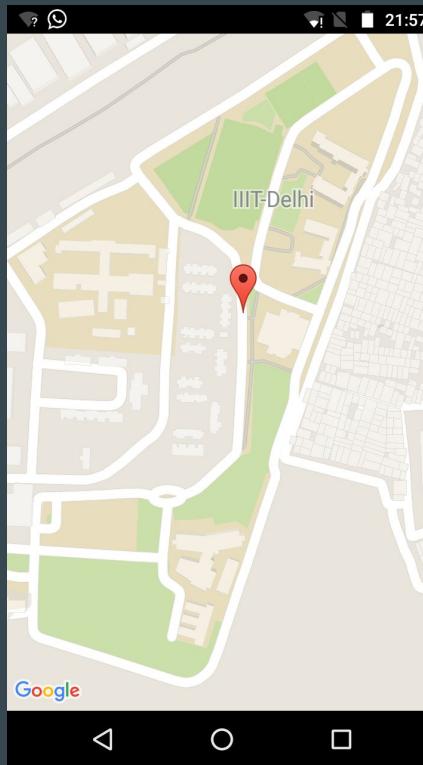
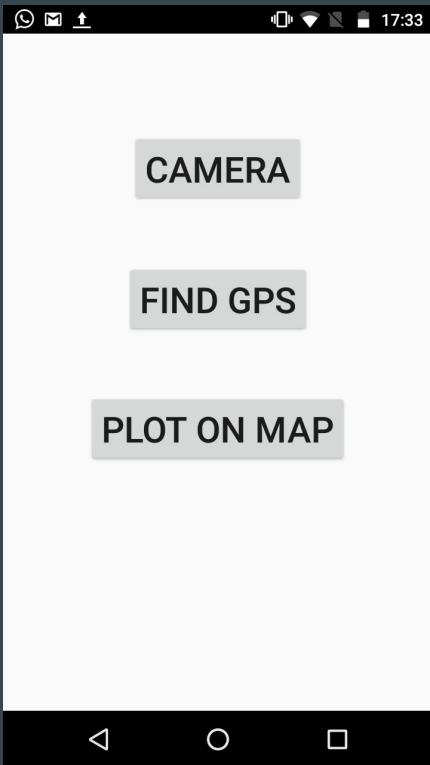
A. 1-vs-All SVM

- a. Two Classes : Positive/Negative
- b. Features : Bag of SIFT (Dimensions : 50/100/1500)
- c. Training Set : Maintained training ratio of 1:4 (Positive:Negative Images)

B. Multi Class SVM

- a. Ten Classes : 4,3,3 faces for Student Center, Library Building and Boys Hostel, respectively
- b. Features : Bag of Sift

Server and Mobile Application



Retrofit 2.0

A type safe REST Client for Android & Java



Data Collection and Dataset

Two Spots were chosen along a straight line from each of the (open) faces of the buildings (5m and 10m away)

From every spot 5 images were taken. The data of each of these images was later parsed in a 6 dimensional tuple.

A typical tuple corresponding to an image looks like :

(NE,28.54,77.27,1455617230886.jpg,F,1) =

(Direction,Lat,Long,FileName,Face,Building-Index)

Student Center : 300 Images

Library Building : 200 Images

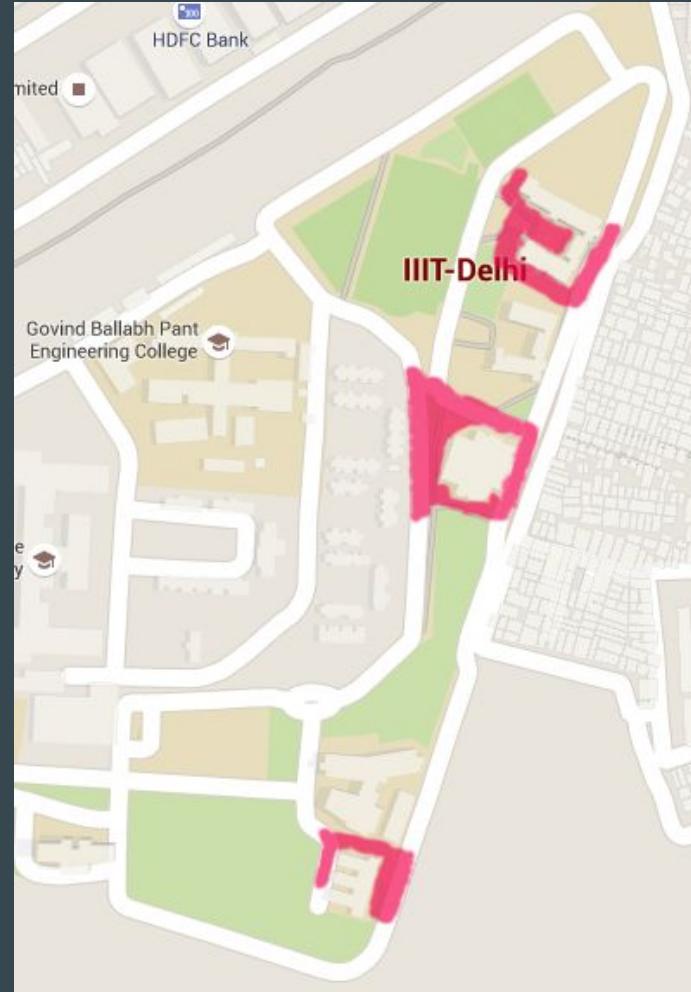
Boys Hostel : 150 Images



Link to [Database](#).

Evaluation Criteria

- A. Location Detection (Classifier Prediction)
 - a. Accuracy = $(TP + TN) / (TP + FP + FN + TN)$
 - b. Sensitivity (TPR) : $TP / (TP + FN)$
 - c. Specificity (TNR) : $TN / (TN + FP)$
- B. GPS Localization
 - a. The error in the GPS measurements by our physical devices was large (~10 m), and hence could not analyse properly



Results

- Maximum Error in GPS estimation: 3m
- Location Recognition :



Face of Building	Accuracy	Sensitivity	Specificity
Back(Facing West) SC	0.73	0.64	0.75
Front (Facing East) SC	0.86	0.79	0.82
Entry Face (Facing South) LB	0.72	0.60	0.88
Back Face (Facing West) LB	0.62	0.54	0.64
Front Face (Facing West) BH	0.81	0.72	0.82
Left Face (Facing North) BH	0.80	0.66	0.83

1500 bags,
1-vs-All Classifiers

Future Works

- Use cross-validation to measure performance rather than the fixed test / train split.
- More accurate GPS estimation using structure of motion techniques
- Add a validation set to tune learning parameters.
- We can try using the various Machine Learning models, like Artificial Neural Networks, Random Decision Forests, to classify the building faces, and compare the results.
- Add spatial information to the features by creating a grid of visual word histograms over the image, as discussed in Beyond Bag Of Words by Lazebnik et al.

Questions ?

Thank You