

Building Identification Using AR Drone

Dhruv Kumar Yadav, 11253, dhruvkr@, CS365A
 Smith Gupta, 11720, smith@

Introduction

In the recent times UAVs have become a major source of active research since they have immense potential in search and rescue operations, sports and autonomous tracking. Many areas where UAVs can be used are devoid of facilities like GPS. In this project we propose a novel method for determining location of an AR Drone in such a GPS denied environment which is achieved using identification of surrounding environment through image classification of surrounding buildings in the terrain of IIT Kanpur.

Procedure

TRAINING

A. Video Segmentation

❖ The video is first segmented into various images at the frame rate of .5 (i.e. a image per 2 seconds of clip) using ffmpeg library.

B. Feature Extraction

❖ Scale Invariant Feature Transform (SIFT) descriptor compute 'key points' and 'descriptors'. SIFT allows to detect the interest points under scale, rotation and noise.

❖ We employed dense sift rather than the sift described by Lowe's algorithm. Dense sift compute descriptor over dense grid of pixels. For category recognition, it has been found that dense sift features give better result than sift features. [1][3][4]

C. Bag of Visual Words

❖ We take the output (feature descriptors) from feature extraction in the first step using SIFT features. Next step is to quantize descriptors into visual vocabulary. K-means clustering algorithm is employed and each cluster is assigned to words to obtain dictionary of k-visual words. There are 128 x K matrix of visual word centers.

❖ Next each image represented as histogram of these visual words. These histograms are used for classification.[4]

D. Multi-Class SVM

❖ Positive and negative vectors for the classifier are the histograms, one from each training image. The classifier here learns all the patterns and regularities in the input vectors.

❖ One-Vs-One classification is chosen which improves the speed

❖ With successful training, classifier is able to classify a test sample into correct class with high accuracy.[3][4]

TESTING

Given a test image, it's test score is calculated through which probabilities are assigned to each class and maximum out of them is selected. Method proposed in [2] is used which assigns probability to each class for given test score in the following manner,

$$\Pr(w_i | (X = x)) = \frac{1}{\sum_{j=1, j \neq i}^K \Pr(w_j | w_{ij} \wedge (X = x)) - (K - 2)}$$

$\Pr(w_i | w_{ij} \wedge (X = x))$ refers to probability associated with class w_i in One vs. One scenario incorporating classes w_i and w_j . This is calculated via the Bayes rule,

$$\Pr(w_i | w_{ij} \wedge (X = x)) = \frac{p(x | w_i) \Pr(w_i)}{p(x | w_i) \Pr(w_i) + p(x | w_j) \Pr(w_j)}$$

where $p(X = x | w_i)$ and $p(X=x | w_j)$ are class-conditional densities that are governed by probability distribution of classes in each of the One vs. One classification pairs which we calculate using distribution of training images over the range of test score. Normalization is done at various steps to make probabilities of classes exhaustive.

Conclusions

The project demonstrated that Building Identification is feasible using an quadcopter equipped with an on-board camera. This is achieved by bag of Visual Words for representing each image in the form of a visual word and these words are then clustered using K-means algorithm. Finally One vs. One Binary SVM Classification algorithm is used for learning the classifier.

Fututre Work

SkyCall

An autonomous flying quadcopter and a personal tour guide build at MIT Senseable City Lab that uses GPS to locate the user and flies to them. It flies at a walking speed and leads the user to its destination. The guide also provide interesting information about the sights they pass.

Compare our current results with Categorization through vocabulary tree

Using video stream for classification

Creating a classification algorithm that takes into account the current image along with previous images from video stream.

Extend to more buildings at IIT Kanpur

Experimental Results

Following table shows Total Training Information:

Training Video	Total Images	Front Images	Back Images
H.R. Kadim CSE Dept.	460	224	236
IME	213	176	137
Library	333	189	144
Total	1006		

Accuracy is Modelled as $P(A) = \frac{\text{Number of test images satisfying A}}{\text{Total Test Image}}$

Test Results are shown below

Test Video	Building Identification Accuracy with front test video	Front Side Detection Accuracy	Building Identification Accuracy with back test video	Back Detection Accuracy Method 1
H.R. Kadim CSE Dept.	100	81.81	59.52	45.23
IME	88.37	93.02	84.21	31.55

Following Figure shows the two of he many instances Test Image computed from our code. Image 1 has score of 1.56 positive when comparing IME with Library and image 2 has score of 1.74 when comparing Kadim with Library showing good distinction between Kadim or IME with Library



The A R Drone is not able to distinguish between the two sides of IME which is expected as the two sides two quite similar. Still some accuracy is achieved due to dataset incorporating different surrounding.

References

- [1] Distinctive Image Features from Scale-Invariant Keypoints DAVID G. LOWE 2004
- [2] D. Price, S. Knerr, L. Personnaz, and G. Dreyfus. Pairwise neural network classifiers with probabilistic outputs. In G. Tesauro, D. Touretzky, and T. Leen, The MIT Press, 1995.
- [3] Moranduzzo, Thomas, and Farid Melgani. "A SIFT-SVM method for detecting cars in UAV images", 2012 IEEE International Geoscience and Remote Sensing Symposium, 2012
- [4] Andrea Vedaldi and Andrew Zisserman Image Classification Practical, 2011