Detecting and Localizing Features Using Multi-Class Classification

Amitesh Sah

Electrical and Computer Engineering

NY, USA

aks629@nyu.edu

Neel Dey

Electrical and Computer Engineering

NY, USA

nsd291@nyu.edu

Shashank Garg

Electrical and Computer Engineering

NY, USA

sg4437@nyu.edu

*Abstract*— The project proposes using a Machine Learning based approach to the Computer Vision challenge of detecting and localizing features within an image and categorizing the objects found in that image using a multi-class classification approach. We will use Bag of words special approach to solve this problem which will include the strategies such as Feature Extraction, Code word Dictionary Generation, Training the classifier and discriminative classification, Hierarchical Segmentation.

Keywords—machine learning, SVM, Image localization, Classification, Detection

# Introduction

The project addresses various problems in the field of Image classification like Image classification, localization, detection and segmentation. We will explain each problem one by one below. Image classification problem is the classification of images based on their visual content. It differentiates between images having different objects. For example, the problem of image classification in Figure 1 will be to determine whether the image contains a cat or not? Localization on other hand is to determine if the object present, where in the image is that object present. The red bounding box in Figure 1 shows the location of cat in the image. Image detection in our context will mean to determine the number of objects and their location in an image. Segmentation problem is the accurate partitioning of an image, depicting different objects in an image as shown in Instance Segmentation part of Figure 1.

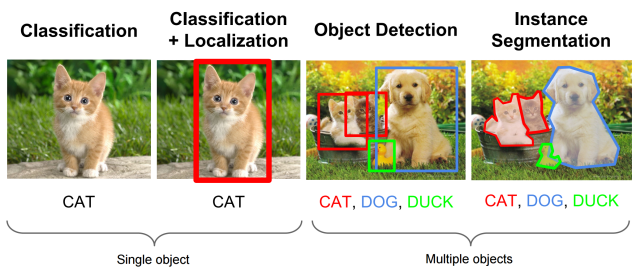


Figure 1: Introduction of Image Classification and Localization

# Motivation

The computer vision industry currently is one of the most growing technology due to advancements in hardware like GPU coupled with parallel processing etc. This has led to various applications being evolved in recent times that deal with various aspects of image and video processing. Traditionally, the algorithms used for these purposes were not accurate due to limitation of their computational cost and lack of data available. Due to availability of open source data has given researchers in the computer vision domain enough resources to come up with new and efficient algorithms to address the problems. One of such problem in Image classification and localization. Image classification is currently used in various industries like medical to detect cancer, security for facial recognition, military for aerial monitoring, image search applications like google image search etc. Coupling the powerful approach of Image processing with the machine learning can give even more powerful algorithms that can help to serve the above applications better.

# Related Work

(Add a reference to each paper you consulted. When you refer to reference #1 in your paper, for example, use this notation: [1] )

# data

We are sampling the ImageNet database of 15 Million Image URLs. However, for a non deep-learning approach, only a few hundred images from each category are sufficient to train our SVM based classifier. A portion of each dataset will also be used for validatiopn and testing the performance of our algorithm.

# Algorithms used

The following explains the baseline algorithm which will be used to address the problems mentioned above.

### Feature Extraction and Hierarchical Segmentation

An image can be broken down into a set of patches. Once these patches are normalized, a descriptor for the patch can be generated. These descriptors can be SIFT, SURF, Harris, etc.

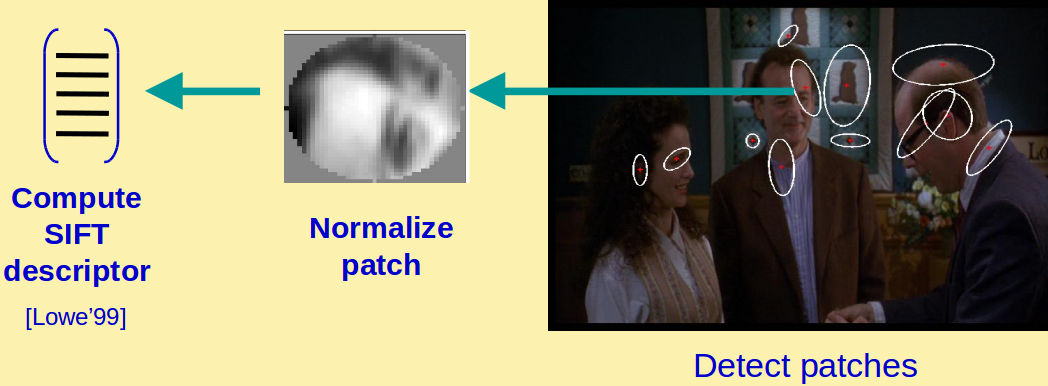


Figure 2: Feature Extraction. Credit: Josef Sivic

### Codeword Dictionary Generation

Once we have multiple descriptors from the images, we use K-means clustering to cluster our descriptors/codewords into groups to generate a histogram of the codewords [6], [7].

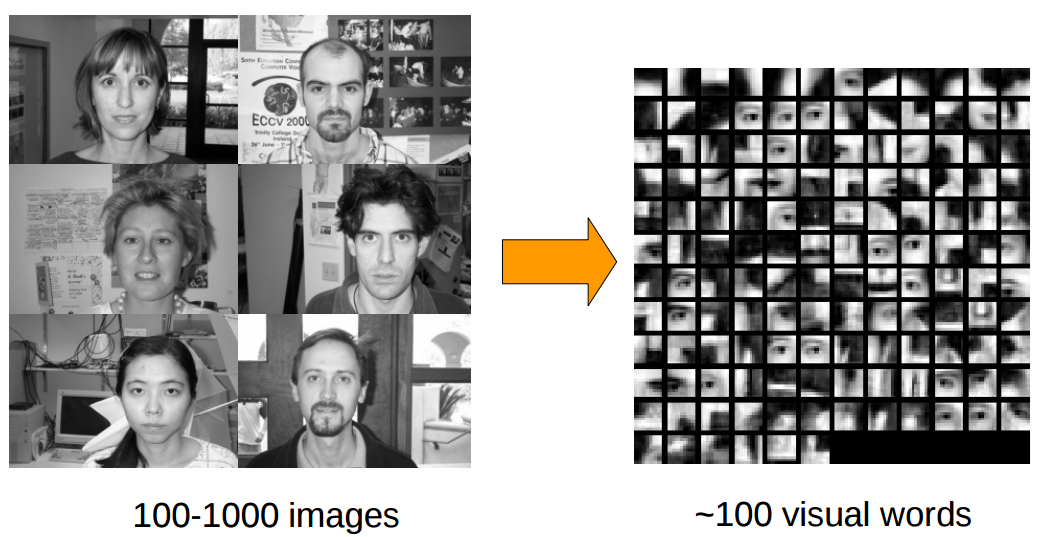


Figure 3

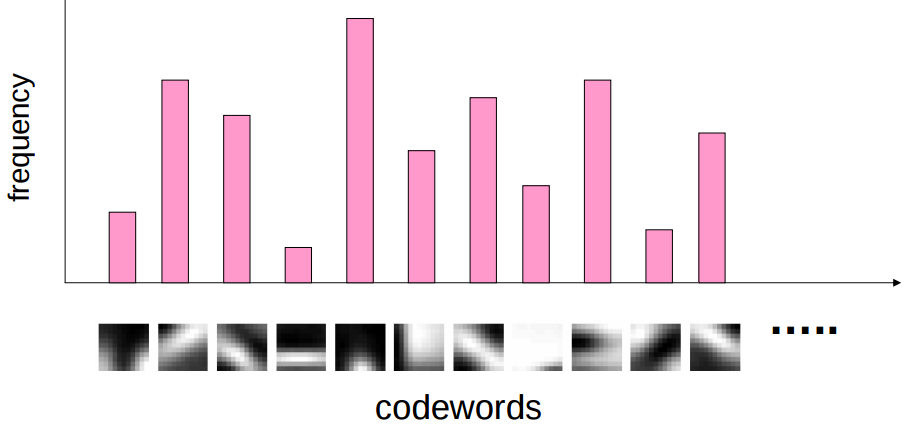


Figure 4: Histogram/Codeword Dictionary Generation

### Training the Classifier and Discriminative Classification

We consider each training image to be a vector which is the Histogram generated previously. We will take a one vs. all Support Vector Machine approach to the classification problem. Other popular approaches to this problem include Naive Bayes, pLAS, etc. and the performance of each will be evaluated with respect to each other.

Since visual words may tend to be highly correlated with each other, it is also best to use a nonlinear kernel when implementing SVM [1] as linear separability within a hyperplane may not exist.

### Hierarchical Segmentation

This form of feature extraction would work well for images with one obvious central object. However, that is rarely the case in real images. For this, we take a hierarchical segmentation approach [5].

We first extract the descriptors from the image. We then segment the image into several pyramid levels of our choosing. The higher the levels, the better the localization. At each level, the image descriptors are segmented. This develops a system where each pixel values votes for belonging to a certain category from lower levels more than it values votes from a higher level. At the end the category with the maximum number of votes wins.

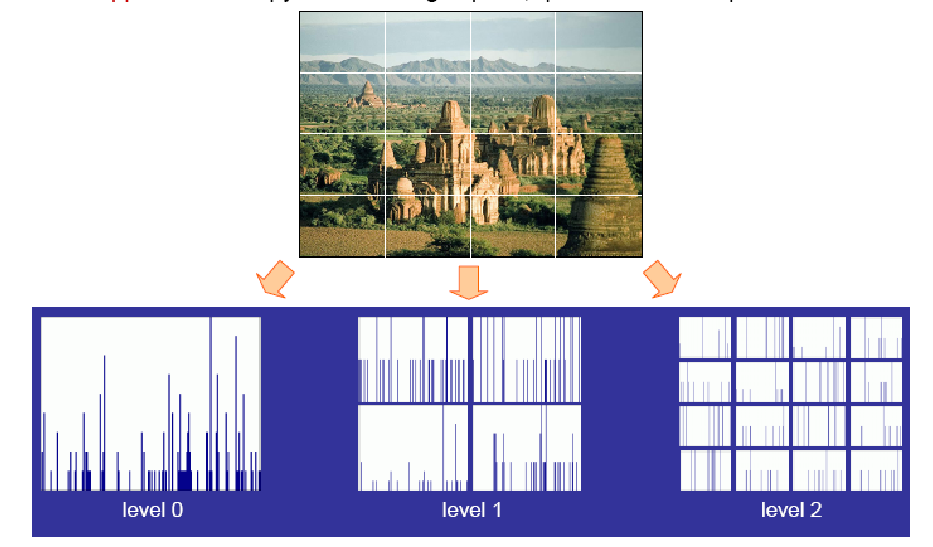


Figure 5: Hierarchical Descriptor Segmentation. Credit: S. Lazebnik

### Data Extraction and Problems Faced

We are using python to extract the data from ImageNet Fall 2011 database. The database provided by ImageNet was not clean and we came across a lot of exceptions and errors which had to be taken care to collect proper and clan data. Some of the urls for images in the database were no longer a valid link to the image. Another problem was that some urls were automatically redirected to different webpages that didn’t contain desired image. Other than that, some of the links although were working were not valid images but a normal html page. We initially were using MATLAB to extract data but due to the exceptions mentioned above we found Python a more convenient and efficient tool to achieve the requirements. The Python script that we wrote takes care of all the mentioned exceptions and errors and we are now able to collect the clean data properly arranged for using in algorithm implementation.

# Code

For the first phase, we have successfully completed the data accumulation from the ImageNet database handling all the exceptions and errors in the urls of the images using a python script available to view on the GitHub link mentioned below. The rest of the algorithm will be implemented on MATLAB.

# Future Work

Future work will include implementing the above-mentioned algorithms on the data collected from the sources mentioned above.

# Conclusion

The data (around 100 images) for 1000 different categories has been successfully collected and is ready to be used in the project.

# REFERENCES

[1]. C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 1998.

[2]. G. Dorko and C. Schmid. Object class recognition using discriminative local features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004.

[3]. L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. In Workshop on Generative-Model Based Vision, 2004.

[4]. J. Sivic, B. Russell, A. Efros, A. Zisserman, B. Freeman, Discovering Objects and their Location in Images, ICCV 2005.

[5]. S. Lazebnik, C. Schmid, and J. Ponce, Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2006.

[6]. L. Fei-Fei, A. Torralba, and R. Fergus, A tutorial on Recognizing and Learning Object Categories: Year 2007.

[7]. J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, Local Features and Kernels for Classifcation of Texture and Object Categories: A Comprehensive Study, IJCV 2007.