

# Enhanced Vocabulary Trees for Real-Time Object Recognition in Image and Video Streams

## Team Members:

- Sugam Jaiswal
- Josiah Zacharias
- Arsheya Raj

May 01, 2023

## Overview

The paper which we chose is called 'Scalable Recognition with a Vocabulary Tree'. Arsheya, Sugam, and Josiah will explore alternative feature detectors like SIFT, SURF, ORB, and others, as well as various tree construction techniques to build a Vocabulary decision tree using the descriptors and K-means clustering to optimize performance in real-time object recognition in image and video streams and real-world applications.

## Goals

1. **Object Detection:** We want to detect the objects for a large dataset of images and videos in the scale of millions. We would like to achieve a better performance than current state of the art object recognition techniques (like deep learning-based approaches) in terms of time and accuracy.
2. **Build the Successful and optimized Vocabulary Tree:** We will construct the Vocabulary tree using the hierarchical K-means clustering algorithm for our initial stage of the project. We can try different clustering algorithms too for constructing the tree.
3. **Make our project as SaaS:** We can construct a web application containing our project and we will deploy it on the web. This is our hopeful goal, to be done at the end once the body of testing and performance is completed.

## Input and Output

The input can be an image/video from a huge dataset of images and video/video streams. Some of the input datasets we might use will be [ImageNet](#), [BDD100K\(UC Berkeley "Deep Drive"\)](#) and [Microsoft COCO](#). The output will be the recognized objects in the images or video annotated by bounding boxes indicating the type and location of the detected objects.

## Evaluation of the Project

We'll be comparing the accuracy and time of the proposed recognition system with the current state-of-the-art object recognition techniques. The level of consistency in the results across different datasets will also be one of the evaluation criterias of our project, including quantitative evaluation against benchmark datasets, including [PASCAL](#) or [Microsoft COCO](#).

Additionally, we will assess the practicality of our approach in real-time applications, such as autonomous vehicles or ..., which will involve qualitative evaluations by testing our system on video streams, simulating these scenarios, and gathering feedback.

## Timeline

Week	Tasks	Notes
May 1 - May 7	Literature review, selection and implementation of alternative detectors and tree construction methods	Project Proposal due May 1
May 8 - May 14	Implementation and testing in OpenCV	
May 15 - May 21	Integration of the enhanced vocabulary trees into a real-time object recognition system	Design proposal due May 19th
May 22 - May 28	Evaluation of the enhanced vocabulary trees and comparison with alternative methods, presentation preparation	Paper presentations 22/24/31
May 29 - June 5	Finalizing the system, extending into SAAS, report writing, and code clean-up	Demo due on June 5
June 7	Submit the Final version of the code and the Project Report	

## Public Source Code

We will use the OpenCV library for implementing the vocabulary tree and object recognition system. Additionally, we plan to utilize the code and ideas provided in the GitHub repository "[Scalable Recognition with a Vocabulary Tree](#)" as a starting point.

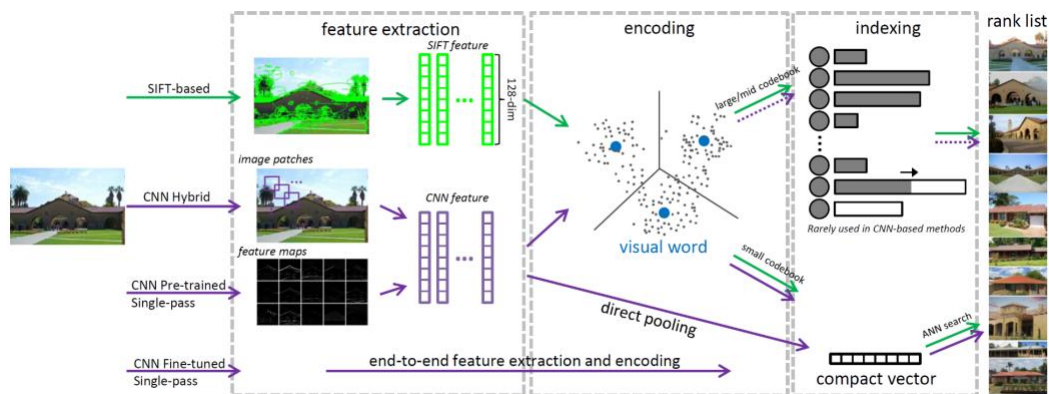


Fig 1: Scalable Recognition with a Vocabulary Tree Extended Pipeline

## Annotated Bibliography

Nister, D., & Stewenius, H. (2006). Scalable Recognition with a Vocabulary Tree. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 2, pp. 2161-2168). IEEE.

[https://inst.eecs.berkeley.edu/~cs294-6/fa06/papers/nister\\_stewenius\\_cvpr2006.pdf](https://inst.eecs.berkeley.edu/~cs294-6/fa06/papers/nister_stewenius_cvpr2006.pdf)

The paper is aimed at solving the problem of scalable object recognition by proposing a scheme to identify and label the objects in images from a large set of labels. Object recognition is one of the most fundamental problems in the domain of computer vision and it has its own challenges like complexity and variability in images. Additionally, an extremely large number of objects to be recognized severely degrades performance of object recognition methods and requires large-scale object recognition techniques. The paper solves this problem by proposing a novel approach for object recognition to handle many objects accurately, like a bag-of-words(features). To do so, the paper presents a hierarchical vocabulary tree constructed by recursively applying k-means clustering on a large set of features to index and match the local features of an image. The paper also uses an inverted file system that stores the image IDs and feature weights for each leaf node allowing fast retrieval of relevant images.