Exploring Computer Vision Techniques: Stop Sign Detection using Cascade Classifiers, Object Detection with Visual Bag of Words, and Visual Odometry

University of New Hampshire

Computer Department Symposium

Shahaab Shokouhi

Stop Sign Detection

Stop sign detection is an important task in computer vision, with applications in various fields such as autonomous driving, surveillance systems, and robotics. In this project, we explore the use of cascade classifiers for detecting stop signs in real-time images and videos.

Fig. 1. Traffic Database

Cascade Classifier

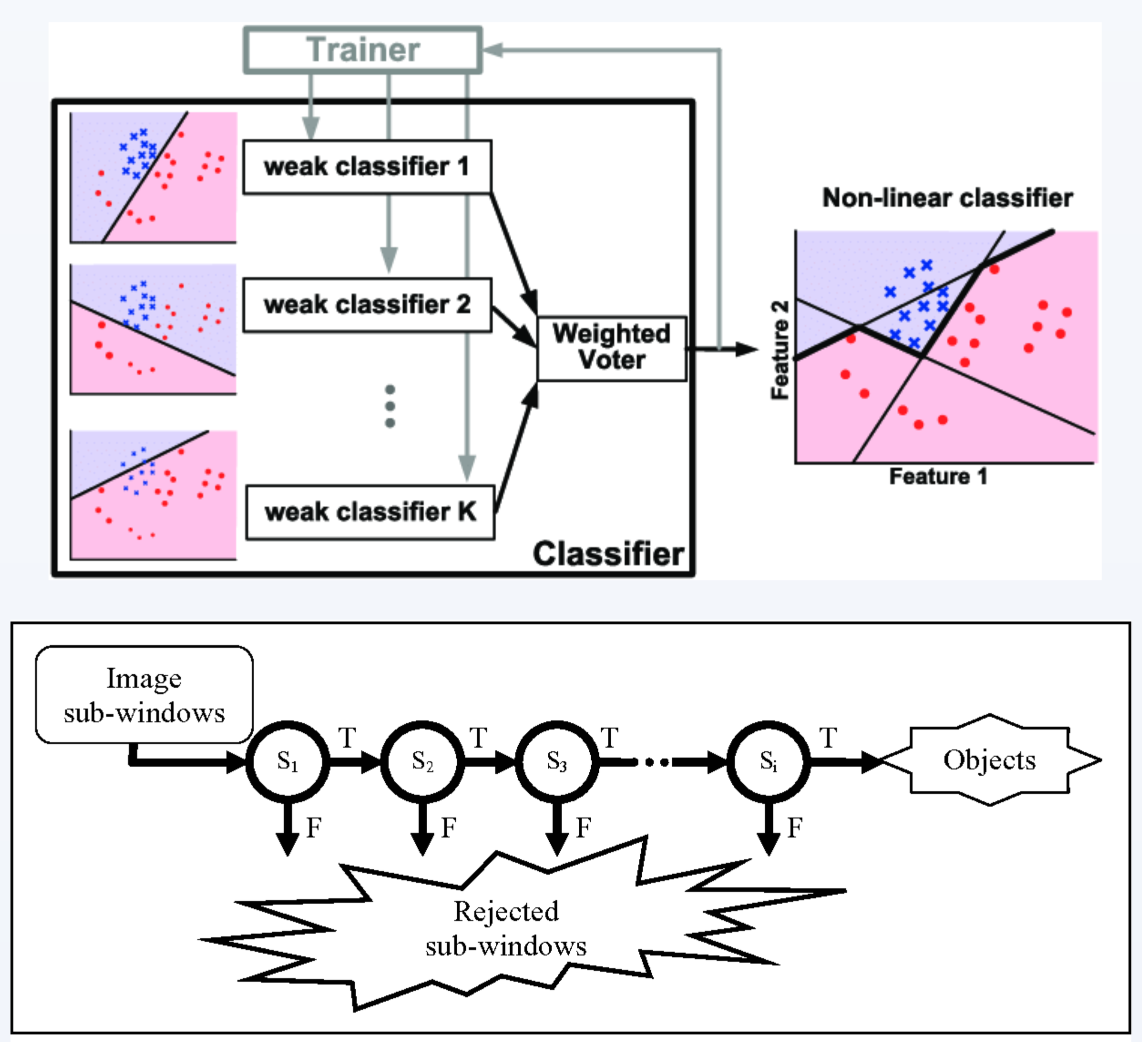

Fig. 2. Configuration of Cascade Classifier

Dataset

The dataset contains 1265 images of 4 distinct types of traffic signs: Traffic Light, Stop, Speed Limit, Crosswalk. About 877 of these images are captured in various Russian cities while the rest are captured in Durham, thanks to Kindrat Beregovyi, a Spring 2022 student who went out of his way and captured those images to enrich the database.

Results

We investigated two parameters to evaluate the accuracy of our classifier: False Alarm Rate (FAR) and accuracy.

$$Accuracy = \frac{TP}{TP + FN}, \quad FAR = \frac{FP}{TN + FP}$$

	True	False
Positive	20	45
Negative	186	6

Based on the above table, the final performance of the classifier is as follows:

$$FAR = 0.19 \ and \ Accuracy = 76\%$$

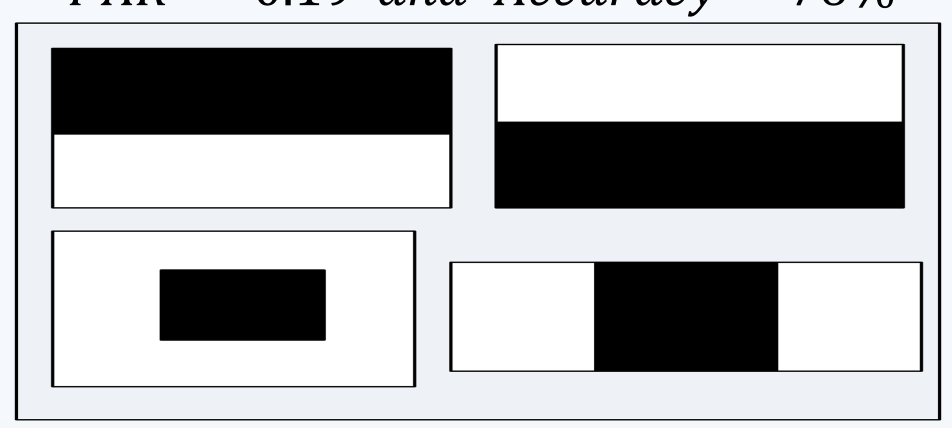

Fig. 3. HAAR features as weak classifiers

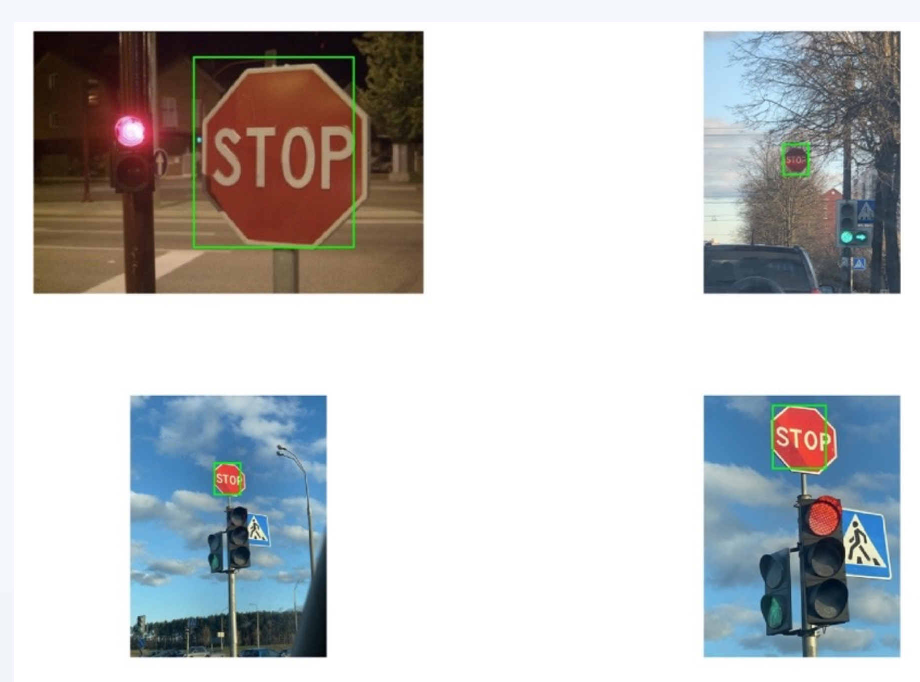

Fig. 4. Detected Stop Sign in four images of the training set

Object Detection with Visual Bag of Words

One approach to object detection is the Visual Bag of Words (BoW) method, which involves extracting local features from the image, such as SIFT or SURF descriptors, and clustering them to form a visual vocabulary.

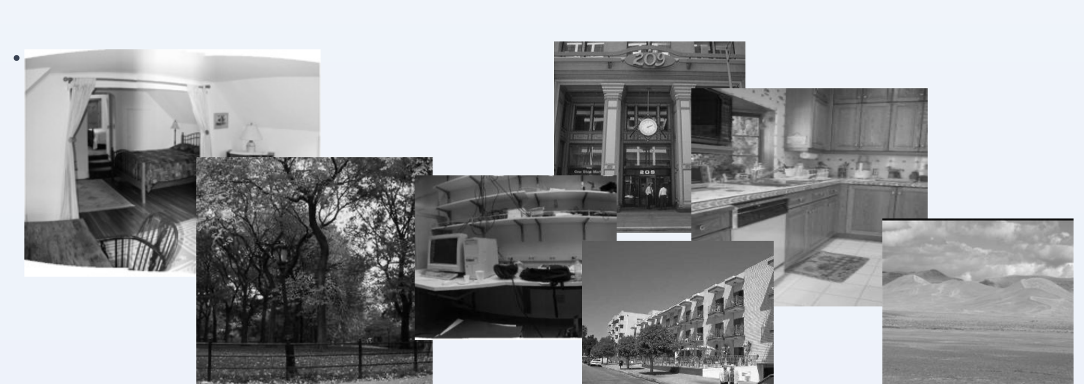

Fig. 5. CVPR2006 Database

Dataset

You will use the 15-scene database introduced in a CVPR2006 paper by Lazebnik et al. The dataset has natural scenes of 15 classes namely, Office, Kitchen, Living room, Bedroom, Store, Industrial, Tall building, Inside city, Street, Highway, Coast, Open country, Mountain, Forest, and Suburb. Each class has 100 training examples and 100 test examples.

Methodology

To start, we tested SIFT Features in combination with SIFT descriptors to create two codebook with different sizes. In addition, we made use of Harris features and HOG descriptors to evaluate their performance as well. We also tested k-mean clustering with two different K values

	$Codebook\ size = 150$		$Codebook\ size = 200$	
Functions	$K = 3$	$K = 5$	$K = 3$	$K = 5$
detectHarrisFeatures extractHOGFeatures	40.46	41.6	40.06	43.46
detectSIFTFeatures extractFeatures	35.4	39.33	35.2	39

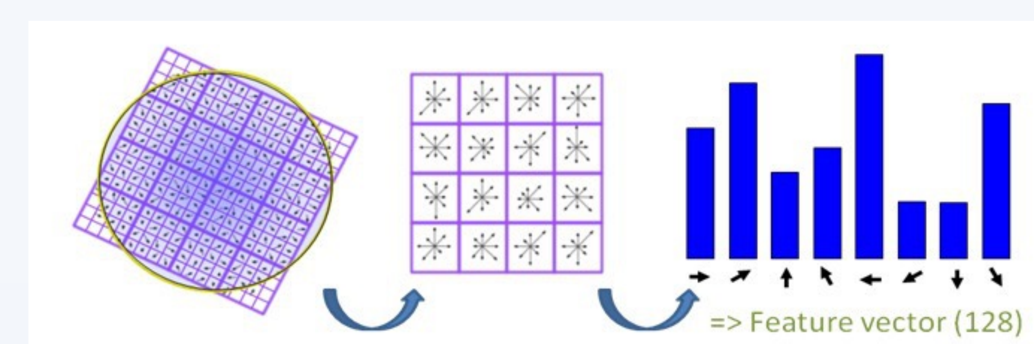

Fig. 6. SIFT Descriptor

Functions	K	Codebook size	Accuracy
detectSIFTFeatures extractHOGFeatures	5	200	46.06

We have also checked the effect of sampling which showed that a sample of 100000 features represents the whole dataset accurate enough. Besides, increasing the K value had impact on the overall accuracy of the algorithm.

Feature work

There are several options to improve the results:

- Using other feature detection and descriptors that are more robust to scale and orientation change
- Optimizing the size of codebook to better represent all categories.

Visual Odometry in GPS-Denied Regions

In many situations, GPS signals may be unreliable or unavailable, making it necessary to use other methods for localization. Visual odometry is a technique in computer vision for estimating the motion of a camera by analyzing the changes in the visual content of successive images captured by the camera.

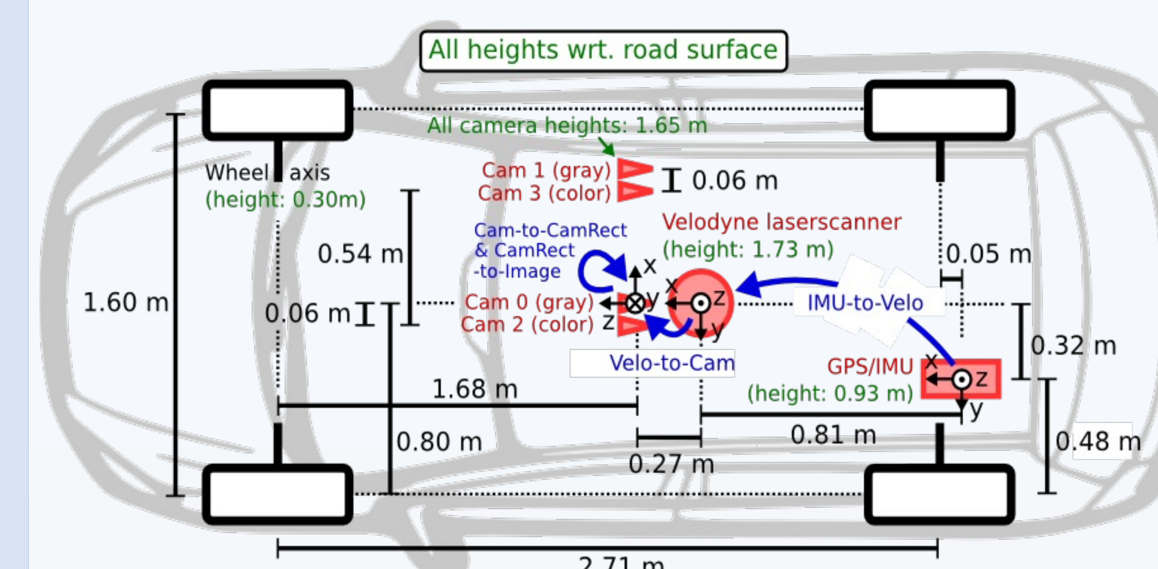

Fig. 7. Top View of Sensors Configuration

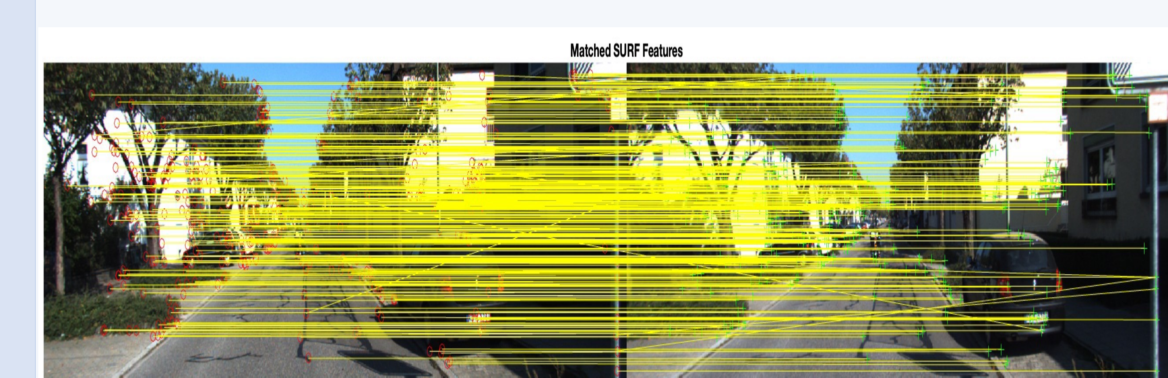

Fig. 8. Feature Detection and Matching

Steps

1. **Feature detection**
2. **Feature matching**
3. **Motion estimation**
4. **Pose update**

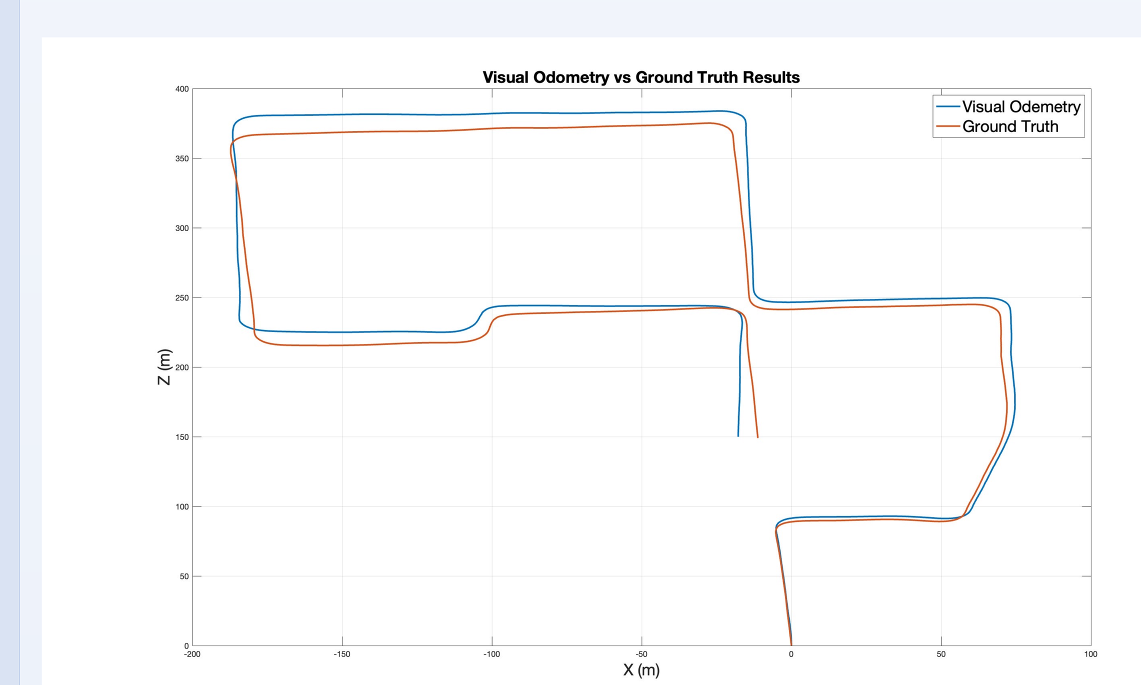

Fig. 9. Visual Odometry vs Ground Truth GPS data