

A large, conical haystack sits in the center of a golden-brown field. A single, thin needle is stuck into the side of the haystack, its point visible. The sky above is filled with large, white, fluffy clouds against a blue background. The overall scene is a classic metaphor for finding something rare in a vast quantity.

# Needle in a Haystack

Carlos Ruiz & Zhengyi Wang

# Content

---

- Introduction
- Problem
- Proposed solution
- Progress
- Planning

} Carlos

} Zhengyi

# Introduction

---

Our project idea comes from an USGS research project:

- Probable locations of a big earthquake's epicenter
- Want to analyze consequences/risk for each possibility
- Knowing the location of energy supplies would be very helpful



# What is the problem?

Electrical companies don't want to provide exact location of their power plants and electrical substations (potential terrorist attacks)

... So what can we do?

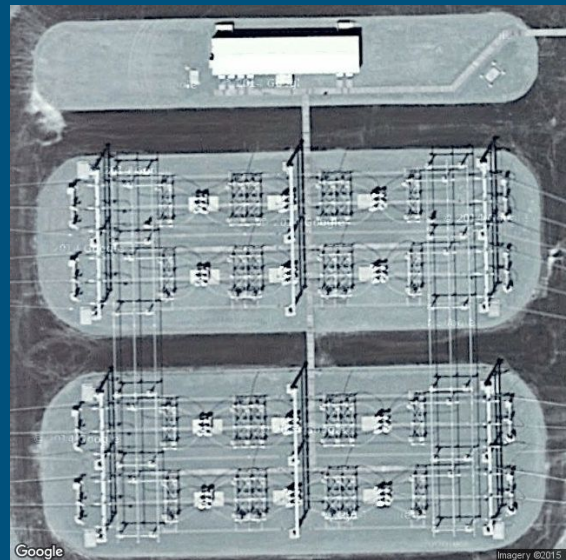
Substation List - LCR

## Local Capacity Area Substation List Based on August 21, 2014

SAP Substation Name	LCR Area	City served	Division	Owner
BRENTWOOD SUB	Greater Bay Area	Bethel Island	DIABLO	PGE
BRITTON SUB	Greater Bay Area	Sunnyvale	DE ANZA	PGE
BRITTON SUB	Greater Bay Area	Santa Clara	DE ANZA	PGE
BROOKSIDE SUB	Greater Bay Area	San Pablo	EAST BAY	PGE
BROOKSIDE SUB	Greater Bay Area	Richmond	EAST BAY	PGE
BRYANT SUB	Greater Bay Area	Orinda	DIABLO	PGE
BURLINGAME SUB	Greater Bay Area	Hillsborough	PENINSULA	PGE
BURLINGAME SUB	Greater Bay Area	Burlingame	PENINSULA	PGE
BURNS SUB	Greater Bay Area	Santa Cruz	CENTRAL COAST	PGE
CALERO SUB	Greater Bay Area	San Jose	SAN JOSE	PGE
CAROLANDS SUB	Greater Bay Area	San Mateo	PENINSULA	PGE
CAROLANDS SUB	Greater Bay Area	Hillsborough	PENINSULA	PGE
CAROLANDS SUB	Greater Bay Area	Burlingame	PENINSULA	PGE
CASTRO SUB	Greater Bay Area	San Francisco	SAN FRANCISCO	PGE
CASTRO VALLEY SUB	Greater Bay Area	Union City	MISSION	PGE
CASTRO VALLEY SUB	Greater Bay Area	Hayward	MISSION	PGE
CASTRO VALLEY SUB	Greater Bay Area	Fairview	MISSION	PGE
CASTRO VALLEY SUB	Greater Bay Area	Cherryland	MISSION	PGE
CASTRO VALLEY SUB	Greater Bay Area	Castro Valley	MISSION	PGE
CASTRO VALLEY SUB	Greater Bay Area	Ashland	MISSION	PGE
CAYETANO SUB	Greater Bay Area	Livermore	MISSION	PGE
CHERRY SUB	Greater Bay Area	San Leandro	MISSION	PGE
CHRISTIE SUB	Greater Bay Area	Hercules	DIABLO	PGE
CLAYTON SUB	Greater Bay Area	Walnut Creek	DIABLO	PGE
CLAYTON SUB	Greater Bay Area	Concord	DIABLO	PGE
CLAYTON SUB	Greater Bay Area	Clayton	DIABLO	PGE
CLAYTON SUB	Greater Bay Area	Alamo	DIABLO	PGE
CONCORD SUB	Greater Bay Area	Concord	DIABLO	PGE
CONTRA COSTA PP SUB	Greater Bay Area	Antioch	DIABLO	nonPGE
CONTRA COSTA SUB	Greater Bay Area	Oakley	DIABLO	PGE
CONTRA COSTA SUB	Greater Bay Area	Brentwood	DIABLO	PGE
CONTRA COSTA SUB	Greater Bay Area	Antioch	DIABLO	PGE
COOLEY LANDING SUB	Greater Bay Area	Palo Alto	PENINSULA	PGE
COCKETT COCKETT SUB	Greater Bay Area	Crockett	EAST BAY	PGE

# Proposed solution

- Machine Learning - recognize substations in Google Maps
  - SVM
  - Naive Bayes
  - Neural Networks
  - Decision Tree
- Feature selection:
  - Color
    - Mean/majority hue, intensity, saturation
    - Entropy
  - Lines
    - Number of lines, sets of 3 (three-phase electric power)
    - Spatial distribution of lines: parallelism/perpendicularity/distance
  - ...



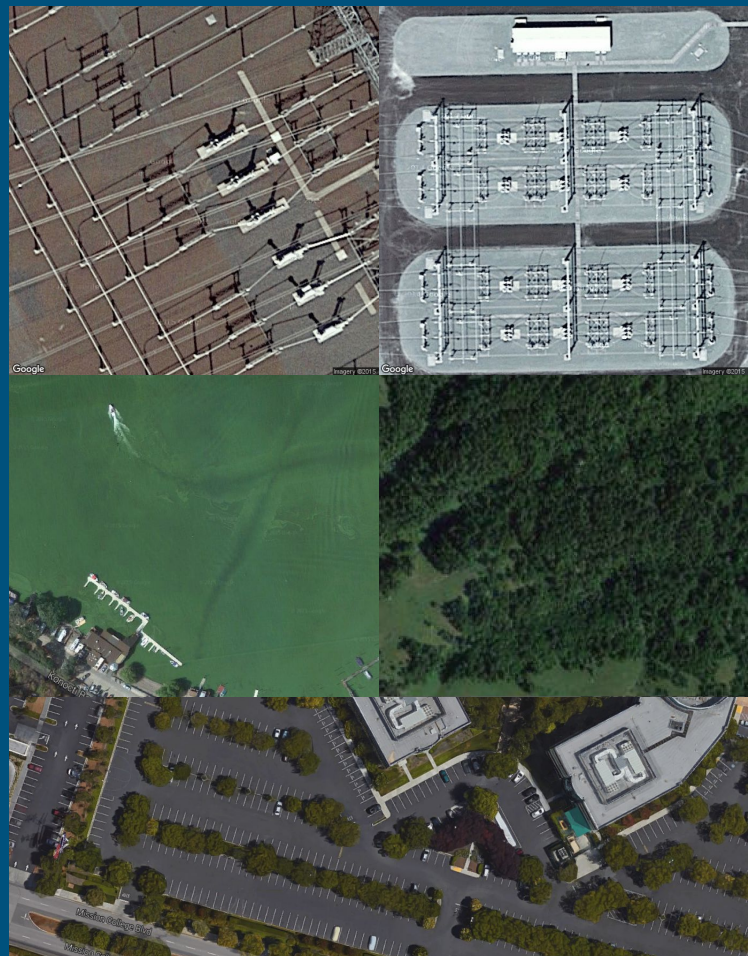


# Progress

---

Preparing our dataset:

- Google Maps API
- Manually find substations +  
Script to download positive samples
  - 25 substations (100-200 images)
  - Most substations are not marked in Google Maps
- Pick negative samples
  - ~250 images
  - Lakes, forests, roads, roofs...



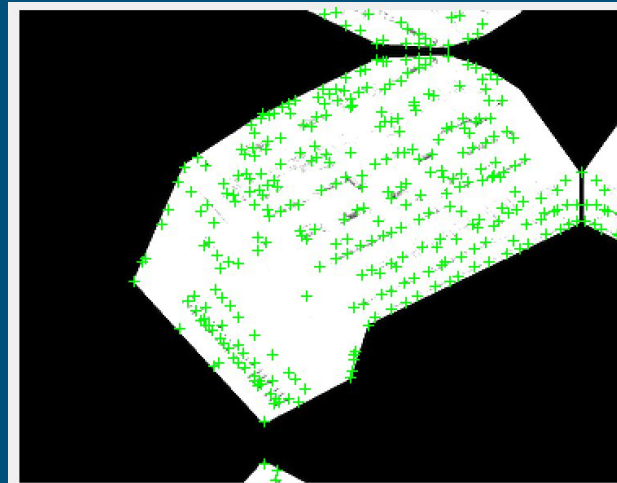
# Progress

---

## 1. Experiment with Matlab Lib

- a. `detectSURFFeatures`
- b. `detectHarrisFeatures`
- c. `detectFASTFeatures`

- 2 sec
- too slow



# Progress

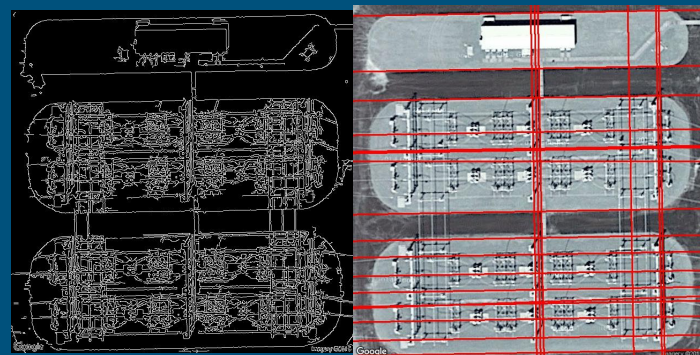
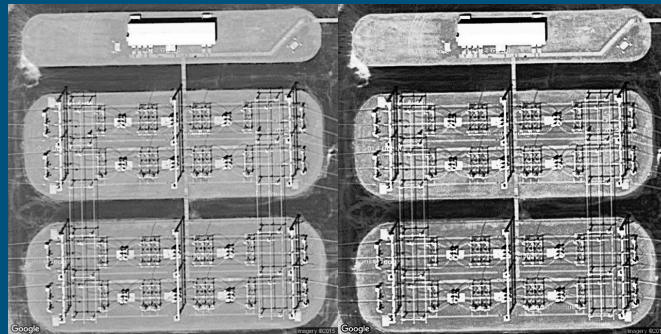
---

- Algorithm Exploring with Python
  - Detect Lines
  - Detect Color



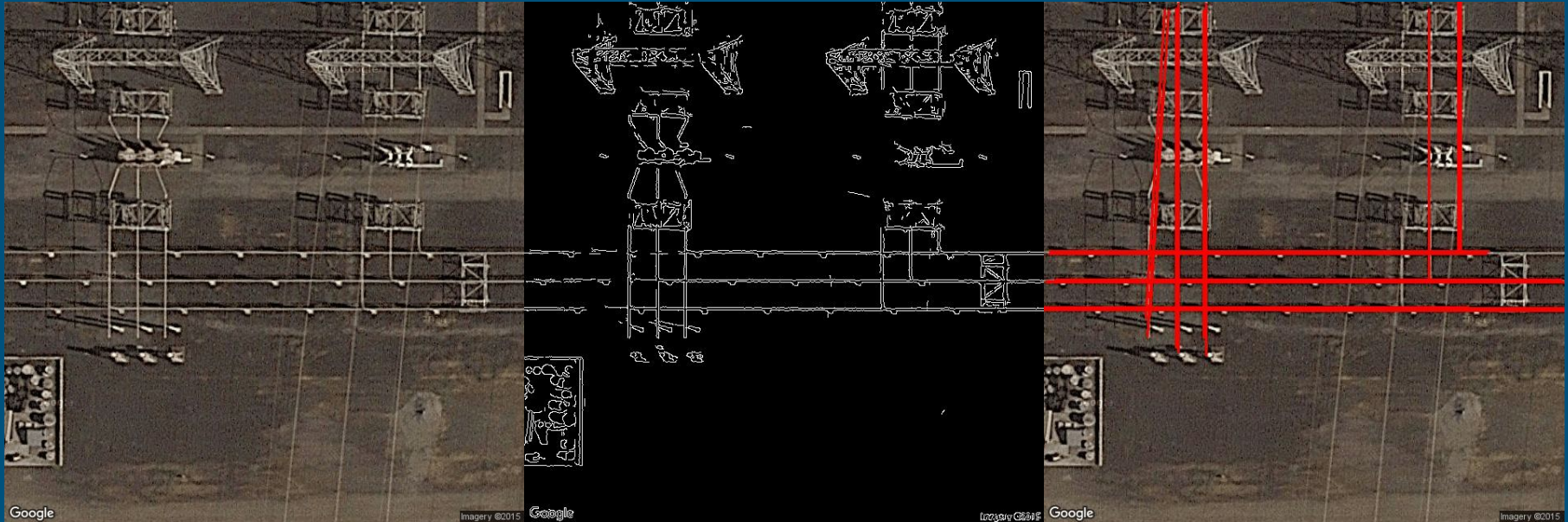
# Progress - Line Detect

- OpenCV-HoughLine
  - improve contrast
  - number of lines
  - sets of 3 parallel lines
  - noise - shadow
  - tuning threshold
- Problem
  - threshold works differently for each picture
  - noise



# Progress - Line Detect

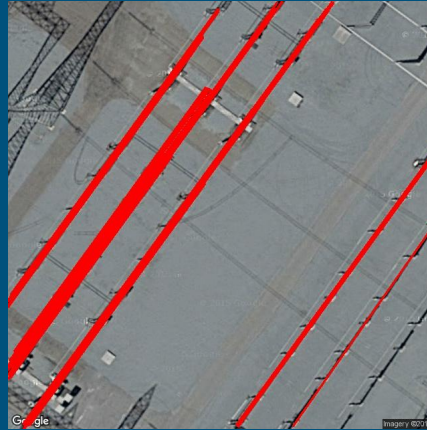
- Edge + Line detection



Original



Line detection



Sample 2

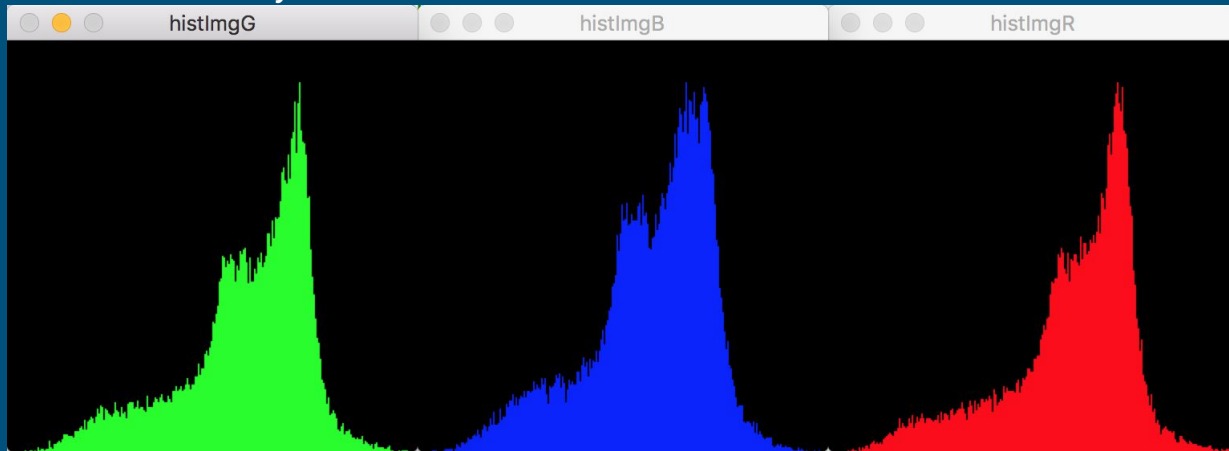


Sample 3

# Progress-Color Detect

---

- Algorithm exploring
  - Color histogram
    - Similarity Metric



# Planning

---

- Explore possible pre-processing steps
  - E.g.: histogram equalization
- Explore which features work better
  - By looking at their distribution over our dataset
- Training with different ML algorithms
  - Compare their test-set accuracy

