Deep Patel
Data Mining
CSE-4334-001

Project Proposal

This project will use two datasets. The first uses UFO sightings data reported to The National UFO Reporting Center, and the other is a collection of images from Caltech called Caltech 101. Together, I will be using them in order to create an easy to use UFO report search tool. Currently, there exists no easy way to search the data NUFORC has on sightings. With my tool, someone who has seen a UFO would be able to look up similar reports with a text query. Additionally, images that relate to the query will be shown. These images will hopefully give context to the report, or maybe even suggest an explanation as to what the UFO really was.

This project will support many features. The first of which will be the search features. The search feature will allow a user to search for UFO sightings that have similar descriptions. This is useful if the user has seen a UFO themselves and wants to confirm if others have seen anything like it before. Additionally, the dataset includes a 'shape' description of the UFO, as well as a location and time. With this data, searching by location, shape, or time should all be possible.

The next feature will be a classifier feature. This feature will ensure that a more detailed search can be made by classifying the UFO descriptions. This will be a powerful feature in that it will be able to group any similar descriptions together even if they are far apart in time, location, and day.

Finally, the image search feature will use the Caltech 101 image dataset in order to provide an image that relates to the users' query. Caltech 101 is an image dataset that has many images belonging to 101 categories. Each category can have anywhere from 40 to 800 images.

A few similar apps do exist, but nothing quite like what is described above. All the similar sites found below lack a search feature. Particularly one that finds similar reports.

- **Metrocosm.com** (http://metrocosm.com/ufo-sightings-map.html)
  - Displays reports on a map according to number of witnesses.
- **Arcgis.com** (https://www.arcgis.com/apps/webappviewer/index.html?id=ddda71d5211f47e782b12f3f8d06246e)
  - Also uses a map in order to display sightings.
- **History.com** (https://www.history.com/shows/project-blue-book/pages/ufo-sightings-location-map)
  - Is another side that uses an interactive map to display UFO sightings.

Links to datasets:

- UFO Reports - http://www.nuforc.org/
  - Csv file version - https://github.com/planetsig/ufo-reports
- Caltech101 images - http://www.vision.caltech.edu/Image_Datasets/Caltech101/#Description
  - Requested Citation - 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*. IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004