Electricity is essential for health and economic prosperity, but 16% of humanity has no access to it. Understanding and assessing electricity access is difficult because:

1. Acquiring reliable data relies on inefficient and often inaccurate surveys [1]
2. Maintaining up-to-date data is costly, with many current cases of incomplete data

We propose a comprehensive method for cheap, quick assessment of power access at a scale that will allow policymakers, researchers and both private and public companies to understand root causes of electricity disparity, and to target areas in need appropriately.

**Overview**

Electricity is essential for health and economic prosperity, but 16% of humanity has no access to it. Understanding and assessing electricity access is difficult because:

1. Acquiring reliable data relies on inefficient and often inaccurate surveys [1]
2. Maintaining up-to-date data is costly, with many current cases of incomplete data

We propose a comprehensive method for cheap, quick assessment of power access at a scale that will allow policymakers, researchers and both private and public companies to understand root causes of electricity disparity, and to target areas in need appropriately.

**Objectives**

We want to develop a method that can efficiently determine from aerial imagery whether or not a village has access to electrical power. This will be in the form of a supervised machine-learning classifier, for which we need a ground-truth dataset for our training and testing data.

Our work this summer has focused on creating two data sets for this purpose: 1) Indian village electrification information 2) US power plants. As this project continues past this summer, a Bass Connections team will use these datasets as training data to create electricity maps and energy infrastructure maps around the world.

**Methods**

A variety of objects in aerial imagery can indicate electricity access. We investigated a long list including but not limited to cars, street lights, substations, and transmission lines. Considering availability and visibility restrictions (e.g., transmission lines are hardly visible even in very high-resolution imagery), we narrowed our focus down to power plants, irrigated fields, and lights-at-night. The three indicators were divided into two datasets due to the distinction and/or similarity in their distribution and detection characteristics.

**Annotation Tool & Power Plant Dataset**

To automate the process of gathering ground truth data, we developed an annotation tool on the Amazon Mechanical Turk platform, with which we created a dataset among 8,503 US power plants because the presence of these were most likely to be visible from public aerial imagery. Features such as NDVI and green index were found to help determine electrification by measuring irriation in a region. Though government-provided groundtruth data can be partially inaccurate and incomplete in some cases, it is still readily available for all Indian states, and so we can employ it as simply a base model to help train our classifier. By building upon this information with additional features from satellite imagery, we can enhance current data by increasing accuracy and predicting electrification for perhaps undocumented villages.

**Products**

1. **Online annotation tool:**
   - [github.com/JayH/AnnotationTool](https://github.com/JayH/AnnotationTool)
   - We created a tool capable of collecting image annotation from Amazon Mechanical Turk. Any images can be used, and any annotations can be collected as long as they can be denoted with a point, line, or polygon. Multiple images can be annotated and multiple annotations can be collected simultaneously. This tool was the source of our annotations for the power plant dataset.

2. **Power plant dataset:**
   - Types: Oil, Gas, Hydro, Solar
   - The power plant dataset consists of four subsets:
     - (a) High-resolution power plant images with accepted annotations (Scope: U.S.)
     - (b) Medium-resolution Landsat images (Scope: global)
     - (c) Annotations labeling the power plant within the image (ground truth for machine learning)
     - (d) Metadata

3. **Indian village dataset:**
   - [https://github.com/SatishSundrija/IndianVillage](https://github.com/SatishSundrija/IndianVillage)
   - Table 1 shows Landsat images (Sept. 2016) and a subset of metrics that are particularly distinguishable for sample villages whose electrification rates come from the map shown in Figure 5. Compared with the unelectrified ones, villages with over 99% electrification rate look brighter in the night; in the last quarter of 2016, they showed greater variation in vegetation coverage with almost identical rainfall statistics.

**Conclusions & Future Work**

We have demonstrated an approach to create datasets that can be used as training and validation data for various machine learning applications. (See Figure 6)

The power plant and irrigation datasets were created because the presence of these were most likely to be visible from public aerial imagery. Features such as NDVI and green index were found to help determine electrification by measuring irrigation in a region. Though government-provided groundtruth data can be partially inaccurate and incomplete in some cases, it is still readily available for all Indian states, and so we can employ it as simply a base model to help train our classifier. By building upon this information with additional features from satellite imagery, we can enhance current data by increasing accuracy and predicting electrification for perhaps undocumented villages.

Both datasets could be improved with higher resolution, multispectral imagery. This would allow for detection of small fields and the inclusion of smaller infrastructure objects such as substations. Higher resolution lights at night imagery would similarly improve the Indian villages dataset.

We also consider expanding the datasets with more categories of electricity indicators and broader spatial coverage.

The two datasets we developed this summer will be inherited by the Bass Connections team in Fall 17 – Spring 18, where researchers continue the work on assessing electricity access in developing countries with statistical and machine learning algorithms.

**References**


