Center for Embedded Networked Sensing

# Imagers as sensors: Correlating plant $CO_2$ uptake with digital visible-light imagery

**Josh Hyman, Eric Graham, Mark Hansen, Deborah Estrin**

Center for Embedded Networked Sensing - http://www.cens.ucla.edu

## Introduction: Use imagers to predict hard-to-measure natural phenomena

### Some phenomena are difficult to measure

- **Existing sensors are hard to use in the field**
  Measuring some biological phenomena require ***bulky*** or ***invasive*** sensors which make deployments difficult to manage and increase experimental error.
- **More creative sensing techniques are required**
  These issues suggest the use of a ***model*** based on other sensing modalities which can provide sufficiently ***accurate prediction*** as a substitute for direct measurement.

### Imagers are an untapped sensing modality

- **Domain knowledge can suggest meaningful ways to process images of a given biological phenomena**
  Biologists know which signals are important for understanding a particular phenomena. This ***domain specific*** knowledge can be codified into a set of ***image features*** which we can compute.
- **Models based on image features turn imagers into first-class biological sensors**
  These models can predict ground-truth that isn't readily apparent from the images, a technique we call ***applied vision***. Unlike general vision, the ground-truth is not easily recognizable by humans.

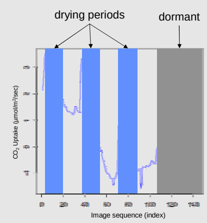
## Problem Description: Estimating $CO_2$ uptake of a drought tolerant moss from images

### Field measurement of $CO_2$ uptake is bulky and invasive

- **What is $CO_2$ uptake?**
  $CO_2$ uptake ($\mu$mol/m$^2$/sec) is the amount of $CO_2$ ***absorbed*** or ***released*** by a plant during photosynthesis. Domain knowledge suggests that the ***greenness*** of the plant should be a good predictor of $CO_2$ uptake.
- **Why is $CO_2$ uptake important?**
  Dense measurement of plant $CO_2$ uptake can be extrapolated to entire forests. Such measurements can be used to refine the model of the ***global carbon cycle***.

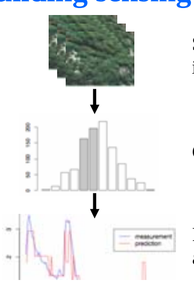
### Experimental Setup

- **Water the moss and cycle light source on and off**
  Simulate a rain event and subsequent days and nights. The moss dries as water evaporates; during darkness, the water is redistributed. When dry throughout, the moss becomes dormant.
- **Measure $CO_2$ uptake using spectroscopy**
  Moss samples are kept in a controlled environment (left top), intake and exhaust air's $CO_2$ contents is measured (left bottom). Labeled graph of $CO_2$ uptake from one drying cycle shown to the right.




## Proposed Solution: Build statistical models from ground truth data gathered in the lab

### Building sensing imagers



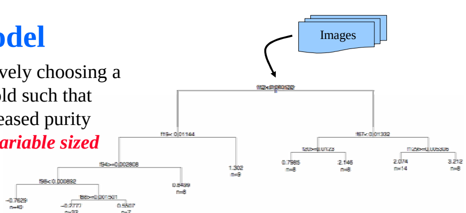
Start with images taken of a biological event stored in a database

Compute domain relevant image feature set

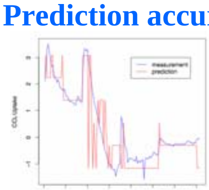
Find the most correlated image features and generate a model from this set to predict the biological event

### Regression based model

Build a ***regression tree*** by recursively choosing a feature and corresponding threshold such that examples in child notes have increased purity (resulting tree on right). Creates ***variable sized*** pseudo-bands based on the data.




### Prediction accuracy

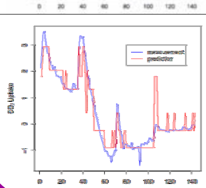


**Classification based model**

| | |
|---|---|
| Lab sensor measurement error | 0.10 μmol |
| Acceptable model error | 0.50 μmol |
| Model's RMS error | 0.74 μmol |

Explanation of error

- too few classes to cover all drying states
- many classes had too little training data
- classes of fixed size, doesn't reflect reality



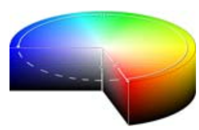
**Regression based model**

| | |
|---|---|
| Lab sensor measurement error | 0.10 μmol |
| Acceptable model error | 0.50 μmol |
| Model's RMS error | 0.49 μmol |

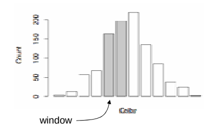
Better prediction

- nine (9) "bands" chosen intelligently
- "bands" are ***data adaptive*** rather than static

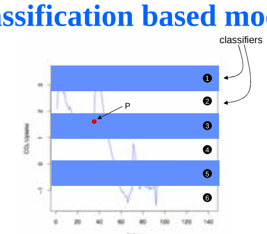
### Extracted Features



- Domain knowledge suggests that color, specifically ***greenness***, is a good predictor
- Compute HSV (Hue, Saturation, Value) ***histogram***
  - more stable than RGB
  - inexpensive to compute
- Compute a set of variable sized ***windows***, grouping similar colors (right bottom)




### Classification based model



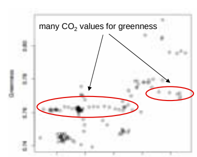

| Classifier | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Response | -2.19 | 1.54 | 3.08 | 2.33 | -1.37 | -3.21 |

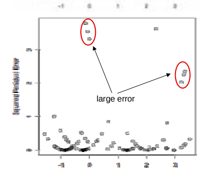
Response is the distance to the hyper-surface separating class K from all other points

- **Training**
  Divide range of values into six ***equal-sized*** bands and train six binary ***SVM*** based classifiers using all generated features.
- **Prediction**
  A point P is in class K if the Kth binary classifier responds most strongly (table to the left). A classified point P is assigned the ***median*** value of class K.

### Error analysis

- Likely locations of high error revealed by rudimentary greenness measure (right top)
- Many different $CO_2$ measurements for a greenness value makes prediction more difficult
- The squared error is ***parabolic-shaped*** because one value is assigned to each "band" (right bottom).




**Biological reason:**

- Same approximate color for different stages of drying
- Moss drying stage dictates $CO_2$ uptake

UCLA – UCR – Caltech – USC – UC Merced