Crowdsourcing-driven bioindication framework

Definition of problem

Originally the #Aircheck challenge mentions the tool for public entry of one's symptoms and combining them with open data sources of environmental factors. But we recall the quote of famous movie character Dr. Gregory House - "Everybody lies" and decided to collect data from honest organisms - plants. In biology science this method called as "bioindication".

Our solution adds yet another puzzle piece to the publicly available open data sources like pollution measures collected from satellites or monitoring networks.

Purpose of document

Assuming the complexity of challenge and importance of accuracy in algorithms and data structures, this document describes general information and plan for creating the community of volunteers and developers for this project. Participation in “NASA Space Apps Challenge” gives us great opportunity to involve more minds and bring this project to life.

Assumptions in this document are based on research of scientific researches listed in the end of document. Participation of authors of mentioned researches is highly appreciable.

Method of bioindication

Method of bioindication is a common and science-proof method for monitoring environmental conditions because state of biological organisms helps to forecast irreversible changes. Any biological organism gives response to the environment changes and analyzing these responses gives us source of valuable information.

One of the most effective source for bioindication are plants, because they can’t migrate to better location and have to fit in the environment around them. Analysing the way how they adapt provides us possibility to register the most earliest stages of degradation.

Role of fluctuating asymmetry (FA) in bioindication

Analysing the fluctuating asymmetry (FA) of the leaves is the method that better fits into crowdsourcing technology of gathering the source data. It requires measurement of specific parts of the leaves and calculating the deviations from biological symmetry. It works for any biological organism that assumes symmetry in its form, but using leaves is more common because of their relative simplicity.

There are several variations of FA method used by different researchers, which makes impossible to combine the results into single picture.

Basic instruments are required to perform FA measurements prior to computer days are ruler and protractor, later were created software products for batch analysis of leaves like LAMINA, LEAF GUI and others mentioned in section "Existing solutions and researches", however they still require a lot of accuracy and thoroughness multiplied on large number of leaves that needs to be measured, which makes these solutions inaccessible for wide variety of users and therefore - smaller input in the global picture (Kozlov, 2015).

Recent IT achievements allow FA measurement instruments implemented using computer vision technologies and high-resolution smartphone camera, which popularity gives great possibilities in collecting large amounts of data for analysis.

Asymmetry of leaves is a relative value, so it doesn't require identifying of actual dimensions of the leaf.

The following calculation method is best suitable for automation of analysis: Baranov, 2014

Consumers and providers of data

In virtue of method principles and in order to collect large amounts of data, our solution uses win-win strategy, where providers of data becoming also a consumers - every new shot (measurement) not only returns analysis information, but also adds new piece to the global picture.

The following groups of users are considered:

1. Scientific researches, who would benefit from combining of bioindication data with other open data sources like air pollution, ground pollution, urbanization, as well as analysing trends of FA in particular geographic locations and time ranges. Their personal accurate data sets, which will be processed by single method provided by this solution, can be helpful for other researches, who currently spread around the world and often limited in communication.
2. Educational institutions have various programs, where children explore surrounding nature. Our free solution can be great assistance to improve quality and range of measurements as well as be the largest source of data for further analysis by scientific researches.
3. Farmers and gardeners, who can track usage of various agronomic methods and tools, choose better locations and estimate quality of growth of specific cultures in
4. Regular people - everybody, who interested in quality of surrounding environment can be the most powerful provider of data - choosing place for picnic or new home are

examples of every-day usage of our solution. However this group of users can provide less accurate results due to lack of time and professional background, so this part of project is planned for later stages, when appropriate filtering and data adjusting algorithms will be defined.

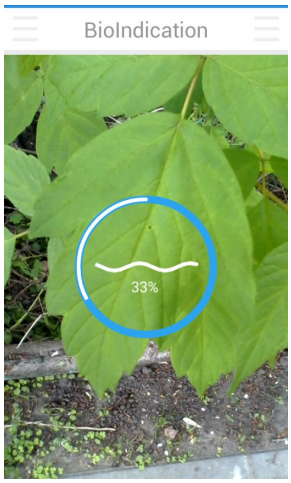
Use cases

Collecting of data

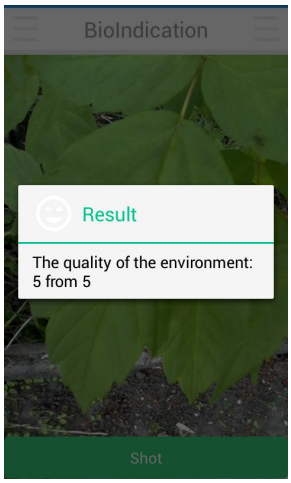
1. User starts application.
2. New submission is initiated.
3. Application indicates status of geolocation and current number of collected samples at concrete cardinal directions.
4. User position camera finder on the leaf and makes a shot.
5. Leaf image is pre-validate and sent to the server.
6. Application suggest to change the cardinal direction of the next shot and continues from step 3.

Application prototype screenshots:

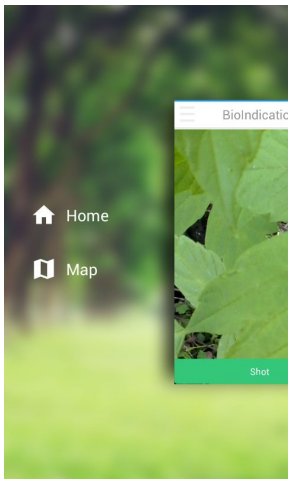
Making shot



Uploading



Actual result



Summary

Solution architecture

Collecting of FA data

Collecting of data (as well as representing) has its specifics for every group of solution users (see chapter “Consumers and providers of data”), therefore this part is implemented in form of open API and several end-user applications, which focus on specific needs of each user group.

API structure

- Registration/Authentication - is necessary for identifying particular sources of information, which helps to filter quality of measurements and perform targeted acts of data collection.

  Every user has its rating based on reviews of quality of his submitted data. The rating can be assigned automatically by the system (based on its ability to perform measurements) as well as other users based on reviews of particular samples (see below).

  Includes the following details:
  - ID/FirstName/LastName
  - Occupation type - scholar, agrarian, consumer.

- Submission of leaf data - each submission contains set of leaf data from particular location point. The more details (samples) comes in submission - the more valuable it is. This value is used for further comparison with other submissions.

  The leaf image is pre-validated on client side by simple image-recognition algorithm to minimize impact of inappropriate (low quality, leaf mechanical damage) data on server pass-through capabilities.

  Submission includes the following details:
  - ID of submission
  - Date/Time, duration of collecting the samples
  - Geographical coordinates
  - Samples (array of items)
    - Leaf image
    - Geographical direction of the photography
    - Possibly data from additional sensors of device

- Deletion of leaf data - the owner of submitted leaf data as well as moderators can delete it
  - ID of submission

- Browse of leaf data - retrieval of leaf data and derived results from data storage by specified conditions.

  Each submission obtains its rating, based on rating of the author of submission, ability of the system to perform automatic measurements and following reviews of submission, made by other users. Users assigned to specific category (occupation) can rate only

submissions of the same category.

Conditions:
- Region of geographical location
- Date/Time period of sampling
- Rating of submission
- FA method identifier

Returned data: in addition to the information collected on submission stage, this method returns the following calculated details of submission:
- Calculated integrated value FA based on algorithm described in Baranov, 2014

- Leaf data rating API - provides the set of methods for submitting/adjusting the rating of submissions (1..5 range) and leaving the feedback.

Existing solutions and researches

John H. Graham, Mattie J. Whitesell, Mark Fleming II, Hagit Hel-Or, Eviatar Nevo and Shmuel Raz, 2015

Fluctuating Asymmetry of Plant Leaves: Batch Processing with LAMINA and Continuous Symmetry Measures
http://www.mdpi.com/2073-8994/7/1/255

Kozlov, 2015

How reproducible are the measurements of leaf fluctuating asymmetry?
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4476141/

Baranov, 2014

Use of Morphogeometric Method for Study Fluctuating Asymmetry in Leaves Tilia cordata under Industrial Pollution
http://www.aensiweb.com/old/aeb/2014/2391-2398.pdf

LAMINA

(Leaf shApe deterMINAtion) is a tool to provide classical measures of leaf shape (blade dimensions) and size (area) as well as measures that indicate asymmetry in leaf shape and measures of herbivory damage (missing leaf area).
https://sourceforge.net/projects/lamina/

LeafProcessor: Measuring and Analyzing Leaf Shape

http://www.iff.fraunhofer.de/en/business-units/biosystems-engineering/research/leafprocessor.html

PlantCV

Plant phenotyping using computer vision
http://plantcv.danforthcenter.org/