CS Capstone Requirements Document

NOVEMBER 1, 2017

Remote Seed Identification

PREPARED FOR

Crop and Soil Science Department, OSU

DAN CURRY

PREPARED BY

Group 12

QUANAH GREEN
ALEX RUEF
ETHAN TAKLA

Abstract

In this document, the hardware and software requirements are defined in detail in order to ensure the timely completion of the remote seed identification project. The project will have a basic minimum set of goals with a set of stretch goals if research goes favorably.

CONTENTS

1 INTRODUCTION

1.1 Purpose

In this document we outline in detail the software and hardware requirements for the Remote Seed Identification project. The requirements and their respective time-lines will be used to guide this project and its development, and additionally to unify the expectations of the project between Daniel Curry of the OSU Seed Lab and Capstone group 12.

1.2 Scope

The software to be developed will have three main components for the remote seed identification; a mobile Android application, a dedicated web server that allows the client to send and receive data from the Jetson TX2, and the seed identification algorithm that will do most of the important processing. The seed identification process is very slow, and this software application will help to significantly reduce seed analysis times and let researchers spend their time on more important work. Due to the fact that the seed industry is quite small, there hasn't been a big effort to automate seed analysis, meaning the end product could have huge implications. Due to the novel nature of this project, a set of basic minimum functionality will be defined, following a set of ultimate stretch goals.

1.3 Definitions

- **Sample:** This refers to a sample of seeds, which will typically be 500 seeds in size. If the project completes its ultimate ambitions, the sample will be 2500 seeds

- **NVIDIA Jetson TX2:** This is a power processor aimed at artificial intelligence and machine learning applications. This is what will run computationally intensive tasks in our back end

- **Open Source Computer Vision Library (OpenCV):** An extensive, highly tested computer vision library optimized to run on NVIDIA processors.

- **Deep Neural Network (DNN):** This refers to one of OpenCV's many features, and is the main component behind the seed identification algorithm.

2 OVERALL DESCRIPTION

2.1 Product Perspective

This projects aims to eliminate manual seed analysis in both Oregon State Seed labs and in the field. The introduction of a simple, quick way to analyze seeds will remove a significant bottleneck that has been frustrating scientists for years. Rather than sending the seed samples to a lab, users will send images of samples to be processed using the Jetson TX2 through a dedicated Android application. Users will be able to view and quantify important results such as how pure the sample is, and if the stretch goals are met should be able to correctly identify and localize seed species in the sample. The application itself will be extremely user-intuitive, enabling even a child to easily use the application.

2.2 Product Functions

As previously mentioned, the remote seed identifier will be split up into three main components: the client-end application, a web server used for sending and receiving data, and the back-end processing unit using.

The client-end android application will be developed as an android application, and will allow the user to easily send pictures of seed samples and get results in minutes. The client application will guide users on how to take a

proper image of a sample, and alert them when their results are ready for analysis. The quality analysis generated for the client will greatly depend on whether the team is able to over-come the issue of low resolution phone and tablet cameras. Additionally, clients will have to make accounts and be approved by an administrator in order to use the seed identification service.

The web server will be constantly running on the Jetson TX2, ready and waiting to accept sample jobs and process them. The web server will only accept jobs that come from registered and approved clients using the app.

Once data is received via the web server on the Jetson TX2, DNN's image classification will be used to analyze seed samples. OpenCV is optimized to run on the Jetsons 256-core processor, so the results will be able to be generated with low latency. Due to the possible performance restrictions associated with low-resolution cameras, the basic minimum requirement for the seed identifier is that it can find rogue seeds in the sample and determine how pure the sample is depending on how many seeds it finds total. As a stretch goal, the algorithm should be able to correctly identify seed species as well, which would be beneficial information to seed analysts.

2.3   User Characteristics

Target users of this tool will primarily be college students and seed analysts, although the intention is for to be usable by anyone that picks up the application. Seed analysts will typically be handling high numbers of samples per day, and need a quick efficient way to view the statistics on their samples. As far as mobile devices, a user will usually be using an iPhone or Android device with camera resolutions as low as 5 megapixels for older devices.

2.4   Constraints

A NVIDIA Jetson TX2 processor will be used to do the image analysis. The power of the Jetson will directly impact the processing time and the format of the input images. Additionally, since many people will be using this app, the phone camera resolution of the devices will vary significantly.The resolution of the input images will directly effect the accuracy of the seed identification software.

2.5   Assumptions and Dependencies

The assumption is made that the average phone camera will have high enough resolution to obtain high accuracy for identifying if a seed is rogue. If a good phone is used, the goal is to also identify the species of the rogue seed as well.

3   SPECIFIC REQUIREMENTS

This section describes in detail the basic minimum set of requirements, as well as stretch goals to be reached as well.

3.1   Functions

- **Minimum requirements**

    1) When taking pictures of the samples, a user should be able to optionally have easy to understand instructions on how to take good photos of the samples.
    2) A user should have regular in-app feedback on the status of the seed processing job so that they know the system is working properly.
    3) A user should be able to save and label analysis results within the application.

4) A user should be able to send analysis results (including saved ones) via email in a PDF format

5) A user should be able to send an image of a 500 seed sample to be processed using the app.

6) A user should receive a notification on their mobile device when their sample has been processed.

7) A user should be able to view an image of the sample which details the location of the rogue seeds so that they can remove them.

8) A user should be able to view the percentage of rogue seeds in the results so that they can get a estimate on the quality of the sample.

- **Ultimate ambitions**

    1) A user should be able to view the species of each rogue seed in the sample so that they can get a better idea of the seeds contaminating the sample.

    2) A user should be able send an image of a 2500 seed sample to be processed.

3.2 Performance and Reliability Requirements

1) Due to the fact that this project is slightly a research one in nature, it would be almost impossible to estimate what the performance requirements should be. A reasonable goal for the basic seed identifier would be a 95 percent accuracy of the purity of the sample.

3.3 Standards Compliance

1) A user analyst should be able to view the confidence the identifier had in the sample, meaning the results can be discarded if it doesn't meet the confidence threshold defined system administrators. This allows organizations to define their own standards for what results they can and cannot use.

3.4 Availability

1) A user should be available 24/7, or as long as Oregon State's networks are working properly, to make the system as flexible as possible for users.

3.5 Security

1) A user should be able to create an analyst account with a password and user name within the mobile application so that they can gain access to the identification service.

2) A user account can be given administrator permissions in order to allow organizations to control who gets to use the service.

3) Administrators are able to approve accounts for usage. This gives a high level of control of who can use the system.

3.6 Maintainability

1) As a stretch goal, a user should be able to upload images of seeds species that currently aren't supported by the app to be trained.

3.7 Portability

1) A user should be able to download the application on an Android device. As a stretch goal, the app will also be supported on iOS.