Progress Report

December 4, 2017

CS461 Fall 2017

Real-time Seed Identification

Prepared for

Oregon State University Crop Science Department

Daniel Curry

Prepared by Omar Elgebaly, Vinayaka Thompson, Xiaoyi Yang

Group 11
Green Team

Omar Elgebaly
Xiaoyi Yang
Vinayaka Thompson

Abstract

This document will give an overview of a project and the design goals. It will discuss the problems we encountered, our weekly updates, and a retrospective indicating what we need to work on for next term.

CONTENTS

1   PROJECT OVERVIEW

1.1   Purpose

In the world today, there is a need to sort seed.

Reasons for sorting seed may include grading for pricing purposes or cleaning for reseeding.

This is an arduous task requiring a researcher to go through thousands of seed one by one under a microscope to determine the seed type and sort it.

The difficulty is compounded when working with small seeds like those produced by grass plants.

Usually, 90% or more of the seed is desired seed.

Thus we are working on a device to reduce the work required to sort seed with a focus on sorting grass seed.

The product we are working on is being designed for the OSU Seed Lab, making the lab-techs working for them our primary audience.

There is, however, a secondary group that this device could be targeted to namely farmers needing to clean small batches of seed for reseeding and other such use cases.

1.2   Scope

The product to be created by our team is a program that is to work with a theoretical set of hardware components to sort seed. We must design our software with our hardware constraints in mind.

This program is designed to take a photo of the seed (in this case Tall Fescue), process it locally, and then send a signal only if an off-type is detected.

The algorithm we make should be able to accurately identify alien seed within a maximum 20% margin of error.

When testing seed, the fully assembled device should be able to process about 14 seeds per second or a batch of 25,000 seeds in 30 minutes.

1.3   Extension Projects

The client discussed a number of extension projects with us that will be implemented as STRETCH goals if we have the time and resources. There was a desire that the device have some sort of user interface which would allow the user to train and test images of other kinds of seeds. As of now, we are only required to discard anything that is not the desired seed, Tall Fescue, but the client mentioned that it would be helpful if there were an interface that will allow a user to train a different type of desired seed. This will be a STRETCH goal and will be implemented should we finish the project

When working on the program, we need to build it in such a way that the following extensions to our product are possible.

The first is the ability for the user to train the device on their own type of seed or be able to switch seed type through a user interface. Since our version will only support the Tall Fescue variant as being the desired seed, this will give the program more robustness by allowing the user to choose a different desired seed type and use this program for more than just identifying seeds differing from Tall Fescue.

The second was categorizing the off-type seed thus giving an estimation of what types of alien seeds are present in the sample. The base version of the program will present an image with various types of seeds and it will output the probability of each seed in the image being Tall Fescue. To categorize an off-type seed, the program would need to be trained on that specific off-type. This means you would have to already know what kind of seed the off-type is and

train a large dataset of that seed. While we cannot necessarily do that because of time constraints, we could support that functionality in a user interface that will allow users to train seeds other than Tall Fescue.

Should we accomplish all the above requirements to the best of our ability with significant time left before the deadline, the client would like us to start on one of the two extension projects.

2   PROGRESS

2.1   Week 1

We met up for the first time as a group this week during class time. In this meeting, we discussed logistics. We sent a schedule we made of all our availability to our client to meet up with him.

2.2   Week 2

We met with the client at 9 am on Friday. In this meeting, he showed us the OSU seed lab. We were introduced to a few seed analysts who gave us a bit of background about what they do. There we learned about OSU's business of sorting out alien seeds from seed packets. We discussed the project with the client and figured out that he wanted us to utilize computer vision to create a program that can identify seeds that differ from Tall Fescue. Additionally, he said he would like our software to work with another mechanical engineering group's capstone project. The plan is to utilize both of our projects to sort seeds in real time. If an alien seed is detected, it will send a signal to the other group's apparatus, which should prompt the apparatus to trash that seed. The client also said the project needs to be quite accurate as mistakes could be costly

2.3   Week 3

This week, we did not meet with our client. However, we completed the problem statement draft and received feedback for it.

2.4   Week 4

We finished and submitted the problem statement final draft. We also sent it to the client and he approved it.

2.5   Week 5

The requirements document rough draft was due this week, so we arranged a meeting with the client to figure out some specifications for the project. We determined two specifications with the client. The first is that our classifier should be able to identify if a seed is Tall Fescue or not with 20% margin of error. In other words, we must achieve a Top-1 score of at least 0.8. The reasoning behind this is because these seed batches get sold to golf courses and landscaping companies. If there are too many alien seeds present in the batch, it could cause weeds and other undesirable plants to sprout. The second specification is that the software should be fast enough to process a batch of 25,000 seeds in 30 minutes. The client did mention he was more flexible on this specification as we are not sure of the limitations of our hardware quite yet. The Jettson TX2 is a powerful processor but it is difficult to estimate how long the testing script would take on such a large batch.

2.6   Week 6

We completed the requirements final draft this week. We sent it to the client for verification and he responded with some helpful feedback. After implementing the necessary changes, the client approved our requirements document.

2.7   Week 7

We needed to complete our technology review rough drafts by the beginning of next week. This means we should get a better idea of what kind of technology we are going to use and what potential alternatives we have. After talking with Kevin, we know for sure that we are going to be provided a high resolution camera and a Jettson TX2. We will have to discuss potential alternatives to these systems as well as potential software we are going to use.

2.8   Week 8

We finished our tech review rough drafts this week. More importantly, our client introduced us to the other group working on the mechanical portion of this project. From them, we figured out that they plan to create a vibrating table with a funnel attachment that will transport the seeds onto the table. The seeds will pass through the image capture area. If the image of those seeds do not get a sufficient Top-1 probability, then we will send a signal to their system indicating that it should be discarded. There are a number of potential issues we discussed here and since we still aren't sure how long it will take an image to be transferred onto the Jettson and tested, we decided that the group should implement functionality for a setting that would control the speed of the vibrating table.

2.9   Week 9

We wrote up our tech review final drafts this week. The hardest part was splitting up this project into parts. We later decided the optimal split will be as follows: data collection, data preprocessing, data training, data testing. Omar will be responsible for preprocessing and training, Vinayaka will be responsible for data collection, and Danny will be responsible for data testing. Each of us is responsible for becoming an expert in the category we are responsible for.

2.10   Week 10

We finished our design doc and progress report this week. We now have a preliminary design for our project. We believe this design will be modified slightly as we progress but it will retain the core features.

3   PROBLEMS

As a group, one of our biggest issues was in communication. We were communicating primarily by email, which made it difficult to for us to all to have the same level of understanding of the project as a whole. It was difficult to organize times to meet up with our busy schedules, which we realized was causing problems. We also realized that our primary form of communication, email, was insufficient. We overcame this problem by setting up a discord channel that will allow us to talk about the project from our homes. We also plan to organize at least one day a week where we all meet to work on the implementation of the project.

On a more technical level, there are various problems that we encountered when it came to designing this project. As a whole, we found it difficult to find resources specifically pertaining to the overarching goal of this project, which

is to identify off-type seeds and sort them out in real time. Since this project is a software component of an overarching project to streamline the process of seed identification, significant testing will have to go into the timing of everything.

While writing the design document, we ran into some significant issues that will still need to be tested for to overcome. The first issue is the way the images will be trained on the classifier. The original plan was to send the seeds on the conveyor belt and have the camera set to take pictures on one second intervals. We figured that if we trained enough images of Tall Fescue seeds, we would be able to get away with only changing the scale and then inputting it into the training model. After some research, we realized this approach may cause a lower accuracy because more objects in the image will cause increased clutter and give more chances for the classifier to produce errors. In addition, we will also have to make sure that each image has the same amount of seeds in it. As such, we believe we will have to either segment the images or send the seeds through the camera one by one, which can be extremely inefficient.

Another issue we ran into while planning the design is figuring out how to integrate it with the final system. Our client expressed desire for our software to work with a setup created by the mechanical engineering group. This is going to be quite difficult and requires extensive testing with the timing. The mechanical engineering group stated they will have multiple settings for the speed of the vibrating table. As was mentioned in the design document, the vibrating table will transport the seeds to the camera where the camera will capture images of it and send it to our Jettson TX2 processor. The Jettson should test the newly captured images against the trained set of Tall Fescue images. If probability that the seed is Tall Fescue is less than 0.8, then we trash that batch of seeds. The design of this part will be incredibly tricky because we do not know how long it will take the pictures to transfer from the camera to the computer. The testing phase will also take some time that we must account for. It is important to note that our Jettson TX2 is supposed to send a signal to the vibrating table to indicate if a batch of seeds is to be discarded. We will have to figure out the timing of all this to be able to automate the testing process with a script. Ideally, the Jettson will receive a batch of images, test them, and then send a discard signal back to the table if the images get an accuracy rate lower than 0.8. The problem is the process is automated, which means that the camera will constantly be taking pictures of new batches. This means that the table will have to either be stopped or slowed down enough to give the classifier time to analyze the image before the next image gets processed. There are a number of potential solutions to this, but it will require significant cooperation from the mechanical engineering group. They would have to include support for modifying the speed of the table, which I believe they said they would.

4 RETROSPECTIVE

Positives	Deltas	Actions
We have solidified our understanding of the project as a group. Over the past 10 weeks, we were able to go from a group of three people with very different views for the project to a group with a unified goal and fairly detailed preliminary design.	In the next term, we will need to be a bit more proactive in meeting our deadlines. Moreover, we will have to make significant strides in improving communication. Our only form of communication this term was through email and we discovered that it was not sufficient. Until we met up as a group at the end of the term to really discuss the design, everyone had significantly different levels of understanding. As such, it created issues in writing the documents. We need to figure out how to use a more immediate form of communication such as voice chat. We also need to figure out a way to use a shared template for writing documents. A problem we had was we all wrote up individual documents and sent them to one person to assemble and turn in. Since everyone used different organization an, the process of combining the document was more difficult than it needed to be.	As of last week, we have created a discord channel for voice chat. We plan on using this channel when writing group documents. This will facilitate a more open communication line between the group, which means we will all have a more unified understanding of the project. We also plan on using a shared latex editor when writing so we can all edit the same template simultaneously and see each other's writing. This will serve to minimize the conflicts in formatting and content that we have encountered and more easily point out errors or concerns in another person's section. Furthermore, we plan on having two 10-15 minute standup meetings a week after class time. In these standups, we plan on discussing our current progress and any stumbling blocks we've come across.