Image Processing in the National Plant Phenomics Centre

Report Name	Progress Report
Author (User Id)	Andrew Tindall (ajt7)
Supervisor (User Id)	Hannah Dee (hmd1)
Module	CS39440
Date	November 18, 2012
Revision	0.1
Status	Draft
Word Count	0

Contents

1 Project Summary

1.1 Phenomics

Before explaining the project, it is first important to understand the field in which it operates - namely that of plant phenomics.

Phenomics is a field of research in biology related to the systematic observation and analysis of "phenomes", or biochemical and physical trait expression; and how their expression changes in result to environmental and genetic factors.

Phenomics is considered a rapidly emerging transdiscipline, requiring expertise in fields including "genetics, molecular biology, cell biology, systems biology, and higher levels of phenotypic expression" alongside wider understanding of mathematical modelling and information sciences. [1]

The discipline has applications across many fields including public health; human genetics; biofuels; food security; and others due to its ability to allow understanding of how factors can affect traits, allowing for more development of more resilient crops, and predispositions to disease.

1.2 Use of Image Processing in Phenomics

Traditionally, methods for phenome observation would involve destructive techniques that remove entire plants or parts, and this results in the need for larger physical space and longer time periods for research. [2]

In recent years there has been a large focus on the development of high-throughput and non-destructive techniques, particularly in the analysis of Arabidopsis - a small flowering plant including Thale Cress which is used as a model organism due to being the first plant to have its genome sequenced in its entirety [4]. Principle among this area is the use of image processing and analysis, which allows for remote screening of multiple traits with minimum disruption to specimens. [2]

LemnaTec are one of the eminent organisations involved in the field of plant phenomics, supporting research institutes around the world through provision of hardware and algorithms. This includes infrastructure at the recently developed National Plant Phenomics Centre (NPPC) at Aberystwyth University which utilises robotic plant handling and automated image capture and analysis to conduct phenomics on entire plant populations in the hopes of identifying plants with increased tolerance to adverse conditions [3]; the work of which provides the basis for this project.

1.3 Project Details

This project seeks to develop a system capable of using multiple algorithms for the processing and analysis of automated image data sets of arabidopsis populations grown in a controlled environment in order to provide phenotype information and analysis about the plants. In doing so, it shall be comparable to, and build upon the the work of the NPPC, and Rosette Tracker [2].

The work has several limitations arising from multiple factors. Foremost is that of data collection. That the project requires the growth of populations of plants introduces inherent time factors into the project, which means datasets are time-limited, and this can hinder early analysis and testing of software. Additionally, variations in the environment for growth, and in terms of imaging, can present discrepancies in the data set, or even render parts of the set unviable for inclusion, such as due to an image being overexposed, or a camera being out of position for several frames of the set. Limitations also exist in terms of what can be achieved through visible-spectrum imaging, which is the broad focus of the project, as not everything

can be observed at the scales and spectrum being used; however it may be possible for the project to utilise further imaging techniques such as IR imaging later into the project subject to the provision of data from the NPPC.

Ultimately, the finished work should output data that provides answers to important questions such as "to what extent does the ede1 mutation affect growth rates and patterns?" and flowering times between different populations. The project may be judged a success should be it capable of providing the prerequisite steps and analysis to reach this stage, as well as data that allows conclusions about phenotypes to be made on these questions and others.

2 Current Progress

3 Planning

Text in here.

Annotated Bibliography

[1] R. Bilder, F. Sabb, T. Cannon, E. London, J. Jentsch, D. S. Parker, R. Poldrack, C. Evans, and N. Freimer, "Phenomics: the systematic study of phenotypes on a genome-wide scale," *Neuroscience*, vol. 164, no. 1, pp. 30 – 42, 2009, linking Genes to Brain Function in Health and Disease. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306452209000487

This paper explores and defines the discpline of phenomics from the perspective of neuroscience, medicine, and in relation to the human genonome. It provides a useful overview of the subject, whilst going into detail of its applications within areas not touched upon in this project.

[2] J. De Vylder, F. Vandenbussche, Y. Hu, W. Philips, and D. Van Der Straeten, "Rosette tracker: An open source image analysis tool for automatic quantification of genotype effects," *Plant Physiology*, vol. 160, no. 3, pp. 1149–1159, 2012. [Online]. Available: http://www.plantphysiol.org/content/160/3/1149.abstract

This paper explores image analysis as a nondestructive method for studying plant growth in Arabidopsis, and presents "Rosette Tracker", which is an open source tool designed to work on both high-throughout and small-scale and low-tech phenomic projects. It looks at reasons for such methods, and outlines clear procedures of a method for plant detection through the use of hue modelling, segmentation, and connected component detection.

[3] Institute of Biological, Environmental, and Rural Sciences; Aberystwyth University, "National plant phenomics centre - accelerating plant improvement." [Online]. Available: http://www.aber.ac.uk/en/media/Example-of-Research---National-Plant-Phenomics-Centre.pdf

An article published by IBERS which outlines the NPPC project, its uses, and the wider context surrounding its work.

[4] The Arabidopsis Genome Iniative, "Analysis of the genome sequence of the flowering plant arabidopsis thaliana," *Nature*, vol. 408, pp. 796–815, 2000. [Online]. Available: http://www.nature.com/nature/journal/v408/n6814/full/408796a0.html

This paper reports the work of the Arabidopsis Genome Initiative in analysis the genome sequence of Arabidopsis, indicating it to be the first time a complete genome sequence of a plant has been presented, and its applications in crop improvement.