Plant Phenotyping with Low Cost Digital Cameras and Image Analytics

Sotirios A Tsaftaris[1] and Christos Noutsos[2]

[1] Department of Electrical Engineering and Computer Science
Northwestern University
Evanston, IL, US
[2] Ecology and Evolution Department
University of Chicago
Chicago, IL, US

Abstract

In this paper we discuss a prototype, easy-to-deploy, and low cost (~ $250) phenotype collection system for growth chambers. Off the shelf digital cameras, wireless transmitters, and PCs are used to store and process the images. A Matlab pipeline is used to fuse multiple images, identify the location of each Arabidopsis plant, segment its leaves, and measure leaf topology and area. Our early findings (unpublished) using this system for correlating genotype to phenotype are very promising. We hope that with future improvements and broad adoption, it will have the same disruptive effects as the first "build your own" microarrayers, which allowed for the explosion of genotyping information. Low cost genotyping and phenotyping will hopefully address some of the environmental, agricultural, and industrial sustainability challenges we are facing today.

1. Introduction

For many years the world scientific community has accepted the reality that the explosive growth in human population, coupled with the inevitable products of modern industry, has profoundly altered our natural environment (Gillet et al. 2008). In recent years, these truths have also entered the

I.N. Athanasiadis et al., *Information Technologies in Environmental Engineering*, Environmental Science and Engineering, DOI 10.1007/978-3-540-88351-7_18,

public awareness; and it is now widely accepted that the changes we have made to the natural environment will have profound implications for human health and welfare (Crosson 1997). Finding ways to achieve true long-term sustainability in industry and agriculture is, in fact, one of the most important challenges confronting us today (Pittock 2003; Hilbert 2001).

The challenges we are facing e.g. habitat/species loss, global climate change (Pittock 2003) are increasingly urgent, and only a sophisticated understanding of the complex biological systems involved in each instance can provide the guidance that conservation scientists and policy makers need. Generating timely, robust, reliable -- and useful -- information about complex biological systems is the key to address many of the world's most pressing policy concerns in diverse areas: public health, human and animal disease, food production, and last but not least ecological conservation.

There exist a few commercially available solutions to plant phenotyping. The ScanAlyzer platform by LemnaTec is a plant phenotyping system focusing in the mature stage of the plant. It has the capability to image plants in a greenhouse by automatically moving plants, placing them on beltways, and positioning them in front of a stereoscopic camera. Proprietary software analyzes the images to extract phenotypic-related information. Although fully developed and tested, this proprietary platform is very costly, requires a large investment in the appropriate infrastructure, and therefore its easy deployment and maintenance are in question. PLENOPSIS, a custom growth chamber phenotyping system, was developed by Optimalog, on contract by the Laboratory of Plant Ecophysiological responses to Environmental Stresses, in Montpellier France (Granier et al. 2005). This proprietary system uses a robotic arm to position an array of sensors on top of a small plant within a growth chamber. As a custom-made proprietary solution there is limited information about its deployment cost. It could be significant, since it requires a particular growth chamber design to accommodate the robotic arm. Furthermore, the imaging software used is not thoroughly presented. Finally, as a proprietary system we could argue that its maintenance might be considered costly.

In this paper we discuss a prototype low cost phenotype collection system for laboratory (growth chambers) based ecosystems. Off the shelf digital cameras, wireless transmitters, and personal computers are used to store and process the images. Our focus is in the simplicity of the acquisition and image processing system. Phenotype analysis and statistical comparisons with actual genotyping information is work in progress.

This paper is organized as follows: first we discuss the importance of phenotyping and the necessity of automated methods. Subsequently, we discuss the hardware and software components of the proposed system. Finally, possible extensions and the conclusion are presented.

2. The Importance of Phenotyping

It is important for our evolution and survival to ensure the existence of many species in an ecosystem. It is necessary to identify all the plant species existing into several areas of the ecosystem collect information about their microhabitat such as temperature, air speed and direction, humidity, light, their exact location. At the same time seeds can be collected from each plant in order to grow them in a control environment, which will resemble the conditions of the natural environment where they were grown and collected. It is important to gather information on each plant's phenotype, which informally translated is how they appear in their natural (or laboratory) environment. By comparing the phenotypes of a plant in a controlled and natural environment it is possible to confirm any differences that were identified in the field sites, and could be contributed not only to different environmental conditions but also to possible plant interactions.

However, the task of generating robust and useful information is impeded by the astonishing complexity of the natural systems involved. The raw number of organisms involved is one aspect of the challenge. For example, intact Illinois mesic prairie can be home to up to 850 species of plants, insects and vertebrates – and an unknown number of microorganisms (McClain 1997). The interdependence of the species at a deep biological level also mandates interdisciplinary research – research that is quite different from the standard academic model.

In order to sustain agriculture which relies on a stable natural environment we need to preserve, maintain, and rebuild complex biological systems (Burel and Baudry 1995). Developing those abilities requires a comprehensive understanding of the systems involved but, sadly, our opportunity to study certain systems is rapidly disappearing. Undisturbed prairie and undisturbed wetland, two distinct ecologies critical to the overall natural health of the Midwestern region, have been denuded and fragmented such that their long-term survival is now in grave doubt. It is necessary to build our understanding of ecological communities at both the molecular (DNA, RNA, proteins, etc.) and macroscopic levels (interactions of species, biodiversity) by characterizing undisturbed systems and experimentally manipulating their component members.

One of the goals in ecological communities is to observe the phenotypes of the organisms in an ecosystem (natural or laboratory) by taking measurements of their appearance (ie., with time-lapse cameras), observing interactions of the organisms (i.e.plants with plants, plants with insects, insects with insects) and store all the data (visual, environmental, genomic) in a dedicated database for further statistical analysis (Edwards and Batley 2004; Lussier and Liu 2007; Boyes et al. 2001). As a first step we check if the observed phenotype is due to interactions with other organisms or not, by growing plants in ideal conditions for them without other organisms

present. Then we try to identify QTLs which are responsible for the observed phenotypes. With the second generation sequencing technologies such as, *454*, *Solexa*, *Illumina*, and SOLiD, the obstacle of sequencing an organism is eliminated (Bonin 2008; Nosil et al. 2008). Thus, it is even possible to do a whole sequencing of populations of the same species, in order to identify areas of the genome, that are responsible for the phenotyping differences. All the generated data can be stored in the same dedicated database for further analysis.

3. System Description

Visual information is critical for the development of a phenotyping system and therefore digital time lapse cameras provide an economical solution to the problem of phenotype collection. Digital images can be easily recorded, transmitted, and stored in a database. However, to amass and analyze the large quantity of information data mining and image analysis algorithms are necessary. The image analysis algorithms are necessary in order to measure all possible characteristics, like the number of leaves on a plant, the length of each leaf, the height of the plant, the existence of any spots in the flower or leaves –an indication of the existence of microorganisms hosted on each plant.

The purpose of this article is to introduce an easy to build and deploy camera-based growth chamber plant phenotyping system, such as the one shown in Fig. 1. Although the proposed system was developed to analyze the early stage leaf development of *Arapidopsis thaliana*, it could be adapted to other model plants.

Fig. 1. Growth chambers with *Arabidopsis theliana* plants. Copyright Borevitz Laboratory at University of Chicago.

The design system requirements are:

1. *Economic hardware*. Economic off the shelf hardware is an enabler of break through and highly adopted technologies.
2. *Easy deployment*. The goal is to build a system that it is easily deployed by non technical personnel such that its adoption could be wide.
3. *Low maintenance*. A system based on off the shelf hardware can be easily maintained since faulty parts can be replaced. This is the premise behind PC based computing clusters that have now dominated the high performance computing market.
4. *Scalable and modular*. Additional cameras can be added in the same chamber monitoring plants from a different angle. More importantly it is critical to have a system that can be programmed and updated easier.
5. *Flash-less operation*. Flash operation will affect the daylight detection mechanism that plants have, and would introduce an exogenous parameter that would severely contaminate the experiments.
6. *Day and night operation*. It is important to acquire images in both conditions to compare the plants behavior in different lighting conditions. Certain mutants may behave differently than others.

In the following section the hardware components and software modules of the proposed system will be discussed.

3.1 Hardware

3.1.1 Image acquisition and transfer

Dedicated digital time lapse cameras, otherwise known as intervalometers, are rather costly, require dedicated support hardware, and have limited expansion capabilities. To overcome this problem we opted for commercially available digital cameras. Specifically, we use CANON PowerShot SD series cameras due to their low cost and satisfactory operating characteristics. Another reason, which will be discussed more below, is the availability of the open source firmware CHDK (http://chdk. wikia.com), which allows for full control of the camera.

To transmit the images to the processing unit, the Eye-Fi (WiFi and storage SD card) is used (http://www.eye.fi/). This card offers a low cost option to adding wireless networking capabilities to any digital camera and provides also onboard adequate flash storage. This solution eliminates additional cables and allows for more compact and weatherproof housing of the camera system. But more importantly allows for the instal-

lation of multiple cameras in one laboratory or chamber, without increasing the cabling infrastructure, which in itself may interfere with the maintenance of the plants in the chamber.

The camera is powered through a USB power adapter that constantly charges the internal battery. This modification was possible through CHDK. The camera and the custom made plexiglass housing (Anderson 2009) were mounted on the chamber using the tripod, shown in Fig. 2. This tripod allows the camera to be mounted to any available grasping location and allows also for easy repositioning of the camera due to the flexible construction of the tripod. At present the camera is mounted on one of the fluorescent lamps used for lighting. In the future it is possible that permanent mounting brackets could be installed in the growth chamber.

For lighting, several fluorescent lamps are used. Some provide ambient light (cool white) while others provide colored light. Green hue lamps are used for nighttime image acquisition, since this light does not trigger the daytime detection mechanism of the plants. Fluorescent lamps are used frequently to regulate lighting in growth chambers. They are controlled by environmental control circuit boards that regulate lighting, humidity, and temperature inside the growth chamber. Since, the intensity of light in both conditions (day and night) was not adequate for flash-less imaging special aperture conditions were necessary. CHDK permits the definition of all camera parameters and it was possible to achieve flash-less photography without any problems and good signal to noise ratio.

A limiting aspect of our setup is the reduced Field of View, which results in some areas being out of focus. One choice would be to find cameras with a better macro lens or use more than one camera. Both solutions increase cost. An alternative solution will be to acquire multiple images with different focus parameters. CHDK is again the enabling factor here, through its Depth of Field Stacking mode. The images are fused together later to provide a fully focused composite image (see next section for details). This solution comes at a bandwidth cost since more images will have to be transmitted, but it is very economical and can accommodate scene and camera placement changes.

Fig. 2. The mounting tripod used. Copyright Joby, Inc.

3.2 Software

3.2.1 CHDK Camera Firmware

The CHDK firmware provides full programming control of the camera via UBASIC scripts that permit the user to control and hardcode exposure and focus, to disable the flash, to define a custom color space (improve green color sampling), optimize compression parameters (reduce color subsampling), and more importantly to acquire time lapse images during the day (ultra intervalometer) and night (long exposure intervalometer). For the operation of the proposed system the ultra intervalometer and long exposure intervalometer scripts were modified. In addition, custom scripts were created to delete the images from the Eye-Fi card (Hause 2008).

3.2.2 Image Processing Module

A background script on the processing computer monitors for the arrival of new images and launches an image analysis module which was written in MATLAB. Automated and unsupervised image analysis is a complicated and computationally intensive task; therefore the image processing module strives in mining phenotype related data from the images.
The image processing pipeline consists of the following four steps:

1. Image fusion to create a single focused image
2. Micropot detection
3. Segment each micropot to find the leaves
4. Morphological analysis on each detected leaf ensemble

At the completion of the image analysis step the original images, the fused image, the location of each micropot, the corresponding segmentation mask, and the measurements of the morphological analysis are formatted appropriately in the XML format and they are stored in a database.

Image fusion: The first part of the module is to fuse the images of different focus to get a fully focused image. There exist a number of algorithms that can do this automatically (Maik et al. 2007; Burt and Kolzynski 1993; De and Chanda 2006; Blum 2005). However, in our current implementation a fixed map was used to combine the images. This map was derived *a priori* using a calibration sheet that had a printed chess patter on its surface, as shown in Fig. 3. The sheet was placed in the chamber and the image acquisition script was launched to acquire the images. The out-of-focus blur was estimated regionally for each image using the method of Kim and Paik (1998). The region with the minimum blur and the corresponding originating image were identified. Using this analysis a map of regions and originating images was used, which guided the selection of

pixels from the original images to form the final focused composite. Further improvements such as the elimination of the calibration phase and a weighted fusion scheme will be discussed in the extensions section.

Micropot detection: Since the micropots are placed onto a plate, which are marked by color tape, the orientation and placement of the plate can be estimated by detecting the color tape markings, as shown in Fig. 4. A rectangular box is drawn within the confines of the tapes. Subsequently, since the number of micropots placed within the plate is known we can identify and crop each micropot. This allows for faster processing and segmentation since each image is rather larger and requires more resources to be segmented in full size. It should be noted that this process only takes place for daylight images since in nighttime images the tape markings are not easily distinguishable.

Leaf segmentation in daylight: We use a color segmentation algorithm to detect and extract the leaves and their location on the plants. The image of each micropot is segmented using color information and dynamic programming clustering with the Wu quantizer (Wu 1992) in the YIQ colorspace (Wikipedia Contributors 2008). Sixteen clusters are used for the segmentation and the cluster belonging to green color is retrieved. If the segmented image does not contain a cluster in the range of green hues the segmented micropot is flagged as inactive and the algorithm continues with the remaining micropot images. Fig. 5 illustrates the results of this process. Figs. 5(a) and (b) show two micropots from different experiments. The image on the right has also a yellow tag that identifies each plant. The next row of images shows the resulting color clusters. It is readily shown that there is a unique "green" cluster that corresponds to the leaves, as illustrated in Figs. 5(e) and 5(f). We have experimented with various cluster parameters and we have seen that the parameters chosen here operate sufficiently for our experimental setup. However, other images or even other plate manufacturers may have a different color distribution. Also, gravels

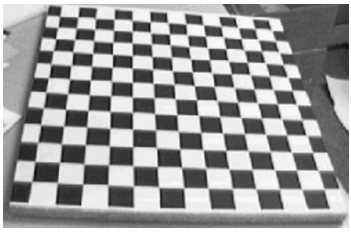


Fig. 3. Square pattern used for out of focus calibration and image fusion.

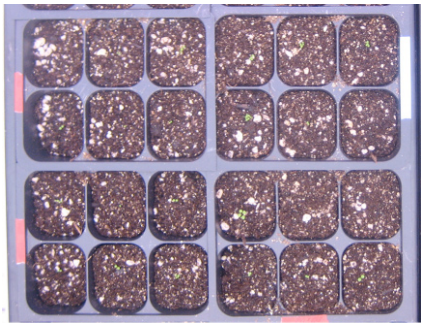


Fig. 4. An image plate with micropots.

in the soil will affect the segmentation result. This can be mitigated by spreading a thin layer of white sand to create a uniform soil background.

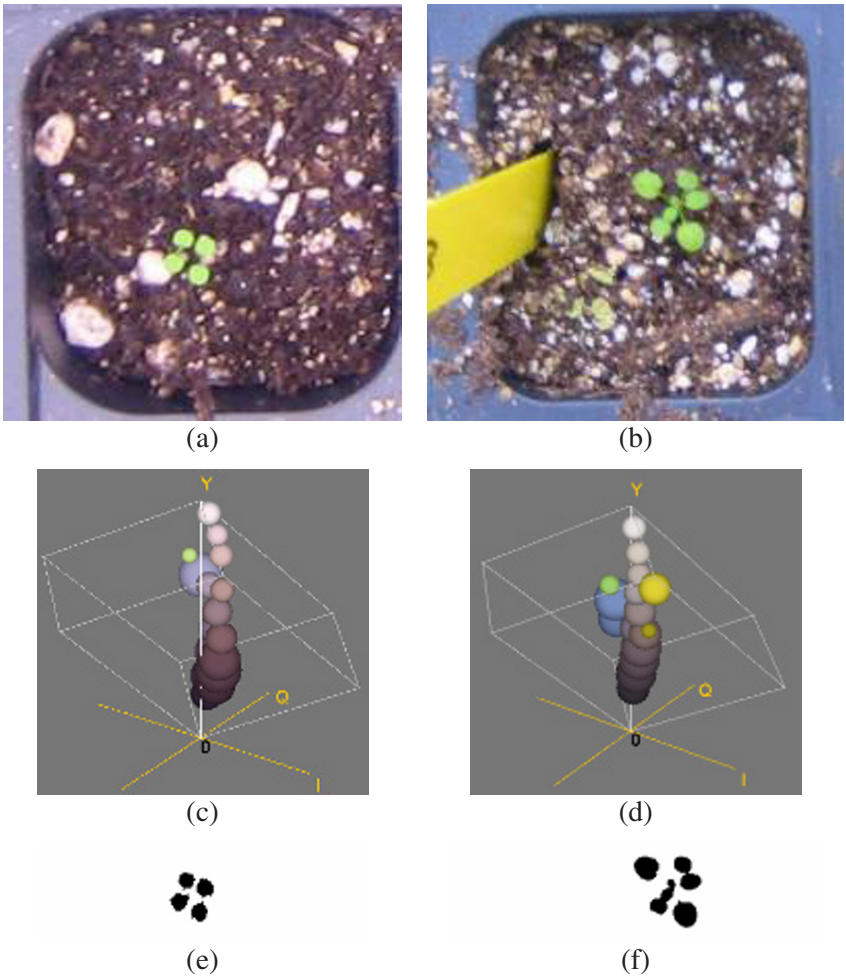


Fig. 5. Segmentation results. (a) and (b) two micropot images; (c) and (d) the 16 cluster Wu YIQ quantization; (e) and (f) binary masks indicating the detection of the leaves for each micropot (cropped around the plant).

Leaf segmentation in nighttime: In nighttime pictures there is limited color information due to the green hue used for illumination. Consequently the algorithm above will not function. Thus the nighttime image is converted to a grayscale only image. The algorithm assumes that the location of each micropot has not changed and it is also assumed that the latest segmentation of the daytime picture is available. For the new night picture, the segmentation mask of the latest daytime image is used to initialize a level set segmentation algorithm (Li et al. 2005). This algorithm operates on the grayscale image and aims in detecting the edges of the leaves.

Morphological analysis: Retrieving the exact topology of the leaves structure requires the exact measurement of the size, orientation, shape, and location of each leaf, which implies the identification of each leaf in the segmented image. This process is accomplished through a series of morphological operations. The binary image is first thinned via erosion to remove connective bridges as the one present in Fig. 5(f). Subsequently, it is connectively labeled using a 4-neighborhood operator. For each label the perimeter is found and an ellipse is fitted using the algorithm of Fitzgibbon et al. (1999). The error between the fitted ellipse and the object perimeter and the eccentricity of the fitted ellipse are used as goodness of fit and overlap measures. Although no further action is taken to resolve the overlapping leaves issues, we believe the measure of error and eccentricity, are adequate for the purpose of this work.

4. Extensions

The proposed system although in its infancy has proved very useful in the data collection of a modern plant biology laboratory. Our hardware cost was bellow $250 (US) and it took about a month of code development for a person familiar with Matlab to develop the image processing pipeline. Statistical analysis and comparison with genotype information is a work in progress and will be presented in an upcoming manuscript. Subsequently, we present certain improvements that warrant mention here so the reader could decide to implement them in advance if he/she were to adapt this system for their use.

We noticed that micropots do not actually stay in the same position. During maintenance and plant care, the researchers tend to move the micropots or even the whole plate. Therefore in the next iteration hard coded (*a priori*) information should be removed. Careful placement of fiducial markers, an example of which is shown in Fig. 6(a), will allow for the complete automation of the system. The markers can be used to identify the location and orientation of the plate. The markers can be poten-

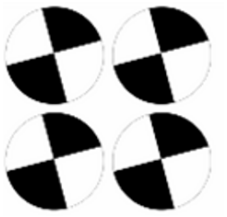
tially printed in specially colored paper to further facilitate their detection. Orientation markers can also be used to allow for the automatic determination of the orientation of the micropots.

Calibration scales such as the one shown in Fig. 6(b), can also be used to assess out of focus blur and potentially eliminate the calibration step (now needed before image fusion can occur). For example the perimeter of the micropot housing can be covered with a thin stripe that is clearly distinguishable from the contents of the micropots. This stripe can be used as a strong edge to determine the out-of-focus blur. Finally, in place of a fixed selection mask used for image fusion, a weighted map can be used to combine information from all available images. There is an abundance of publications in this field with a direct application here (Maik et al. 2007; Burt and Kolzynski 1993; De and Chanda 2006; Blum 2005).

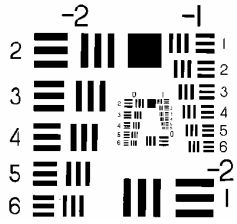
Another possible extension is the implementation of a direct ellipse detection algorithm such as the once presented by Mai et al. (2008). This algorithm will detect ellipses in the micropot images without requiring segmentation, since it relies on edges and their subsequent classification. Furthermore, it can accommodate and distinguish overlapping ellipses. However, since the algorithm detects all possible ellipsoids, only the ellipsoidal shapes of green color must be finally selected.

In terms of software development it will be ideal to develop the algorithm in an open source environment. Matlab despite having a large toolbox library it has a costly license. Python is a possible alternative that is open source and has a large user-developed image processing toolbox. Finally, it is possible to modify the CHDK firmware and add functions that relate directly to this application.

It was expected that some performance will be lost by putting a low cost requirement on the system. The major source for loss of accuracy is the static camera (limited field of view). Using robotic arms or moving cameras improves the captured field of view but increases system cost significantly; it will require actuation devices and digital control to position them.



(a)



(b)

Fig. 6. Fiducial markers for plate detection (a) and scales (b) for size calibration and out-of-focus blur estimation.

5. Conclusions

Although in its infancy we believe that the proposed platform will set the ground for an easy-to-deploy lab-based phenotyping system. Our early findings (unpublished) in using the system in the lab and performing statistical analysis is very promising. We hope that with future improvements and widespread adoption, the proposed system will have the same disruptive effects as the first "build your own" microarray spotting systems, which allowed for the explosion of genotyping information (Derisi 1998). Low cost genotyping and phenotyping combined will hopefully address some of the environmental, agricultural, and industrial sustainability challenges the world is facing.

We believe that we can further improve the usability of this system and make it as universal as possible. Our goal is not to cover every possible plant but provide a scalable and upgradeable system. It is possible that such systems could be implemented to monitor prairies outside the laboratory to classify the flora (Mokhtarian and Abbasi 2004). In the future it is also possible that the service could run in the cloud (outside the laboratory) where independently operating cameras record data and transmit them to processing nodes for phenotyping analysis (Parvin et al. 2002).

Acknowledgements

We would like to thank Prof. Borevitz (Department of Ecology and Evolution at University of Chicago) for providing us with the images and initial funding for this effort. Finally, we should thank Ron Hause, a graduate student from the Committee on Genetics, Genomics, and Systems Biology who did his rotation in Prof. Borevitz's Lab, for his assistance in the development of this project.

References

Nathan P. Gillett, Dáithí A. Stone, Peter A. Stott, Toru Nozawa, Alexey Yu. Karpechko1, Gabriele C. Hegerl, Michael F. Wehner & Philip D. Jones (2008). Attribution of polar warming to human influence. *Nature Geoscience.* 1. 750-754

Pierre Crosson (1997). Impacts of Climate Change on Agriculture. *Climate Issues Brief* 4.

Pittock, B. (2003).Climate change: An Australian Guide to the Science and Potential Impacts, Australian Greenhouse Office.

Granier, et al., (2005). PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist.* 169(3), 623-635.

Hilbert, D.W., Ostendorf, B. and Hopkins, M., (2001).Sensitivity of tropical forests to climate change in the humid tropics of North Queensland. *Austral Ecology*. 26, 590–603.

William E. McClain (1997). Prairie establishment and landscaping. *Nature Heritage Technical Publication*.

Burel, F. and Baudry, J., (1995). Species biodiversity in changing agricultural landscapes: A case study in the Pays d'Auge France. *Agric. Ecosyst. Environ*. 55, pp. 193–200.

Edwards, D. and Batley, J. (2004). Plant bioinformatics: from genome to phenome. *Trends in Biotechnology*, 22(5):232-237.

Lussier, Y. A. and Liu, Y. (2007). Computational approaches to phenotyping: high-throughput phenomics. *Proc Am Thorac Soc*, 4(1):18-25.

Aurelie Bonin (2008). Population Genomics:a new generation of genome scans to bridge the gap with functional genomics. *Molecular ecology*. 17, 3583 – 3584.

Nosil P, Egan, SR, Funk DJ (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: 'isolation by adaptation' and multiple roles for divergent selection. *Evolution*.62,316-336.

Boyes, D. C., Zayed, A. M., Ascenzi, R., Mccaskill, A. J., Hoffman, N. E., Davis, K. R., and Gorlach, J. (2001). Growth stage-based phenotypic analysis of arabidopsis: A model for high throughput functional genomics in plants. *Plant Cell*.13(7):1499-1510.

Tim Anderson (2009). How to Make Your Own Waterproof Camera Enclosure. Online [http://web.media.mit.edu/~tim/pix/waterproofcamera.html]

Ronald J. Hause (2008). Deleting images with interval shooting. Online. [http://chdk.setepontos.com/index.php/topic,2003.0.html]

Maik, V., Cho, D., Shin, J., and Paik, J. (2007). Regularized restoration using image fusion for digital auto-focusing. *Circuits and Systems for Video Technology, IEEE Transactions on*. 17(10):1360-1369.

Burt, P. J. and Kolczynski, R. J. (1993). Enhanced image capture through fusion. *Computer Vision, 1993. Proceedings., Fourth International Conference on*. pp 173-182.

De, I. and Chanda, B. (2006). A simple and efficient algorithm for multifocus image fusion using morphological wavelets. *Signal Process.* 86(5):924-936.

Blum, R. (2005). Robust image fusion using a statistical signal processing approach. *Information Fusion*. 6(2):119-128.

Kim, S. K. and Paik, J. K. (1998). Out-of-focus blur estimation and restoration for digital auto-focusing system. *Electronics Letters*. 34(12):1217-1219.

Wu, X. (1992). Color quantization by dynamic programming and principal analysis. *ACM Trans. Graph.* 11(4):348-372.

Wikipedia contributors (2008) YIQ. Wikipedia, The Free Encyclopedia. Online http://en.wikipedia.org/w/index.php?title=YIQ&oldid=259429685

Li, C., Xu, C., Gui, C., and Fox, M. D. (2005). Level set evolution without re-initialization: a new variational formulation. CVPR 2005. IEEE Computer Society Conference on. 1: 430 – 436.

Fitzgibbon, A., Pilu, M., and Fisher, R. B. (1999). Direct least square fitting of ellipses. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 21(5):476-480.

Mai, F., Hung, Y., Zhong, H., and Sze, W. (2008). A hierarchical approach for fast and robust ellipse extraction. *Pattern Recognition*. 41(8):2512-2524.

Mokhtarian, F. and Abbasi, S. (2004). Matching shapes with self-intersections:application to leaf classification. *Image Processing, IEEE Transactions on*. 13(5):653-661.

Parvin, B., Yang, Q., Fontenay, G., and Barcellos-Hoff, M. H. (2002). Biosig: an imaging bioinformatic system for studying phenomics. *Computer*. 35(7):65-71.

DeRisi, J., Iyer, V, and Brown, P.O. (1998), *The MGuide: A Complete Guide to Building Your Own Microarrayer*. Stanford, CA, Stanford University 1998.