

# **IAP – Ein Informationssystem zur Verarbeitung von Aufnahmen automatisierter Phänotypisierungsanlagen am Beispiel von Gerste (*Hordeum vulgare L.*)**

Alexander Entzian<sup>1</sup>, Kerstin Neumann<sup>2</sup>, Benjamin Kilian<sup>2</sup>, Christian Klukas<sup>1</sup>

<sup>1</sup>AG Bildanalyse, Abt. Molekulare Genetik

<sup>2</sup>AG Genomdiversität, Abt. Genbank

Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK)

Corrensstraße 3

06466 Stadt Seeland OT Gatersleben

{entzian, neumannk, kilian, klukas}@ipk-gatersleben.de

**Abstract:** Hochdurchsatzphänotypisierung ist eine wichtige Grundlage für die biologische Forschung insbesondere im Bereich der Agrarforschung. Zur Bewältigung der dabei entstehenden großen Datenmengen wurde die Integrierte Analyse Plattform „IAP“ entwickelt. Mit diesem Informationssystem sind Bioinformatiker und Biologen in der Lage, effizient Experimente auszuwerten, zu verwalten, zu archivieren, zu visualisieren und statistisch zu interpretieren.

## **1 Motivation und Zielsetzung**

Die Hochdurchsatzanalyse wird in der biologischen Forschung vor allem im Bereich der Microarray-Datenanalyse (DNA, Proteine, ...) und im Bereich der Analyse von Metabolitdaten angewandt. Dieser Trend hat sich nun in den letzten Jahren auch im Bereich der Pflanzenphänotypisierung durchgesetzt [Ha11, Gr06, Wa07]. Auf Grund neuer Entwicklungen [Le12] ist es möglich hunderte von Pflanzen mit vergleichbaren standardisierten Wachstumsbedingungen über ihre gesamte Lebensphase aufzuziehen und nicht-invasiv zu analysieren. Das in dieser Arbeit vorgestellte Informationssystem, die Integrierte Analyse Plattform „IAP“ [KPE12], ist in der Lage sehr große Bilddatenmengen unterschiedlicher Kamerasysteme effizient zu analysieren und zu verarbeiten.

## **2 Aufbau des automatisierten Aufnahmesystems für Gerste**

### **2.1 Kamerasysteme für die Bilderfassung**

Die zu analysierenden Gerstenbilder werden mit dem Hochdurchsatzphänotypisierungssystem von LemnaTec [Le12] mittels drei Bildaufnahmekammern im Bereich des sichtbaren Lichts (1234 x 1624 Pixel), im nahen Infrarot (IR-B, 1450 - 1550 nm; 254 x 320

Pixel) und der Fluoreszenz (420 – 750 nm; 1234 x 1624 Pixel) in 30° Schritten von der Seite und jeweils ein Bild von oben aufgenommen. Bilder des nahen Infrarots geben Rückschlüsse auf den Wasserhaushalt der Pflanze. Die Fluoreszenzbilder basieren auf dem patentierten Dark Reader Setup der Firma Clare Chemical Research [Cl12] und ermöglichen die Erfassung von roten und grünen fluoreszierenden Pflanzenbestandteilen (eGFP, eYFP, RFP, ...) [Eb07].

## 2.2 Überblick über das Gesamtsystem

Das System (Abb. 1 links) besteht aus einer Aufnahmeschleife in der sich die oben beschriebenen Kammern und eine Beladestation befinden (Abb. 1 Mitte). In einem zweiten großen Bereich können bis zu 520 Einzelpflanzen (Abb. 1 rechts) aufwachsen. Beide Bereiche sind mit Förderbändern verbunden, auf welchen die Pflanzen vollautomatisch transportiert werden. Neben der Bildaufnahme erfolgen ebenfalls die Bewässerung und das Wiegen der Pflanzen vollautomatisch nach zuvor festgelegten Zeiten und Werten.



Abbildung 1: Automatisiertes Aufnahmesystem für Gerste; Anlagenüberblick (links); Beladestation und Bildaufnahmekammern (Mitte); Gewächshausbereich (rechts)

## 3 Integrierte Analyse Plattform – IAP

Die integrierte Analyse Plattform (IAP) [KPE12] ist ein System zur Integration verschiedener Datendomänen (Analyseergebnisse, Bilddaten, Klimadaten, ...). Es wurden drei Benutzeroberflächen entwickelt. Eine Web-Oberfläche für das Daten-Browsing, eine Java-Swing-Oberfläche für Interaktionen mit den Daten und ein Kommandozeilen-Client für die Bildanalyse auf einem Server.

### 3.1 Datenzugriff, -import und Backup

Die erzeugten Bilder werden vom Aufnahmesystem in einer PostgreSQL Datenbank mit Metadaten und weiteren Informationen (ID, Rotationswinkel, ...) verknüpft und sind mittels FTP abrufbar. Um mit IAP verarbeitet werden zu können, werden automatisiert einmal täglich alle notwendigen Informationen in das IAP Datenbanksystem (MongoDB) geladen. Zusätzlich, zu den vorhandenen Metadaten, können weitere Informationen wie zum Beispiel Metabolitdaten, Genexpressionsdaten, Klimadaten des Gewächshauses und manuell gemessene Referenzdaten importiert werden. Zur Langzeitarchivierung der

primären Bilddaten und der Analyseergebnisse werden diese in ein hierarchisches Speichermanagement System (Hierarchical Storage Management System, HSM) mit automatischer Bandsicherung kopiert.

### 3.2 Bildanalyse

Die in IAP integrierte Bildanalysepipeline ist in Java implementiert und besteht aus individuell anpassbaren Einzelkomponenten. Als grundlegende Bildverarbeitungsbibliothek wurde ImageJ [SRE12] gewählt. Die Pipeline wird automatisch sequentiell für jede Pflanze vom ersten bis zum letzten Tag für jeden Drehwinkel der Seiten- und Topaufnahmen aufgerufen. In jedem Schritt werden parallel 3 Bilder (ein Bild je Kameratyp) verarbeitet (Abb. 2).

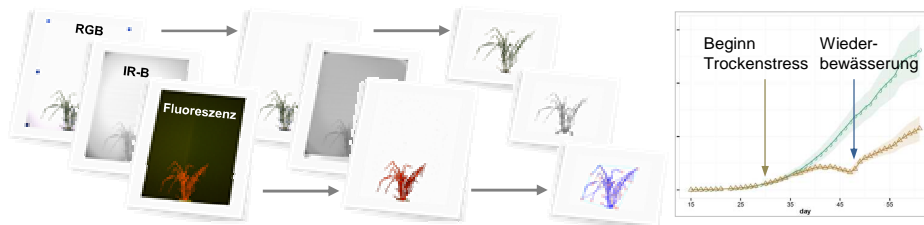


Abbildung 2: Von links nach rechts: Eingangsbild, ausgewählter Zwischenschritt, Ergebnisbild, Verlauf der digitalen Biomasseentwicklung (Mittelwert mit Standardabweichung) von je 156 Gerstenpflanzen unter Trockenstress (braun) und Kontrollbedingungen (grün)

Nachdem die Pflanze vom Hintergrund getrennt wurde, werden anschließend numerische Ergebnisdaten (digitale Biomasse, Höhe, ...) bestimmt. Es ergibt sich folgender Ablauf:

**1. Weiß- und Schwarzabgleich:** Die verwendeten Lichtquellen in den Aufnahmekammern erwärmen sich über den gesamten Tag, auch in Abhängigkeit von der Umgebungstemperatur, was zu veränderten Lichtbedingungen führt. Dadurch kommt es über die Zeit auf den Bildern zu einem veränderten Weiß- (RGB-Kamera) bzw. Schwarzwert (Fluoreszenz-Kamera). Um für jedes Bild eine einheitliche Farbklassifikation zu erhalten, wird deshalb ein Weiß- bzw. Schwarzabgleich durchgeführt. Da die IR-B Kamera systembedingt eine starke Randabschattung aufweist, erfolgt hier ein konzentrischer dynamischer Weißabgleich.

**2. Reduzierung durch Referenzbild und Detektion blauer Markierungen:** Der maximale Wachstumsbereich von Gerste ist mit blauen Markierungen auf der Hintergrundwand der RGB-Kammer markiert. Alles außerhalb dieses Rechtecks wird vom Bild entfernt. Zusätzlich wird vor jedem Durchlauf ein Referenzbild erzeugt („null image“) auf dem jeweils die leere Kammer zu sehen sind. Durch den Vergleich mit diesem Referenzbild kann ebenfalls ein Großteil des Hintergrundes entfernt werden.

**3. Entfernen von kleinen Störobjekten und Anwendung der Fluoreszenzmaske:** Das verbleibende Hintergrundrauschen, sowie die zum Abstützen der Pflanzen notwendigen blauen Metallkörbe werden mittels eines Farbfilters im HSB Farbraum entfernt. An-

schließlich detektiert ein Algorithmus die zusammenhängenden Bildkomponenten und kleineren Störobjekte, die nicht vom Farbfilter erfasst wurden und entfernt diese. Um im IR-B-Bild eine aussagekräftige Trennung zu erreichen, wird zuerst das Verfahren des adaptiven Thresholdings [SP00] angewandt und anschließend das verarbeitete Fluoreszenzbild als Maske verwendet.

**4. Berechnung von Deskriptoren:** Schlussendlich können für alle drei Bildtypen Deskriptoren berechnet werden, beispielsweise digitale Biomasse, Höhe, Breite, die Kompaktheit, die konvexe Hülle und weitere davon abgeleitete Parameter.

### 3.3 Daten-Browsing, statistische Analysen und Report-Generierung

Die Ergebnisse der Bildanalyse können im Web-Client oder der Desktop-Anwendung vom Nutzer betrachtet und durchsucht werden. Zusätzlich werden die Korrelationswerte zwischen experimentellen Daten und manuell erhobenen Deskriptoren (z.B. Frischgewicht und Höhe) ermittelt und angezeigt. Abschließend wird ein Report über alle Ergebnisse erstellt. Dies geschieht in drei Schritten. Als erstes werden die grafisch darzustellenden Daten als CSV-Datei exportiert. Im zweiten Schritt werden die Diagramme mit dem R-Paket ggplot2 [Wi09] ausgegeben, anschließend werden alle Ergebnisse mittels LaTeX zu einer PDF-Report-Datei zusammengefasst.

## Literaturverzeichnis

- [Cl12] Clare Chemical Research, Dark Reader® Technology, Verfügbar: <http://www.clarechemical.com/>, Zugriff: 18.10.2012.
- [Eb07] Eberius M.; LemnaTec Fluorescence Imaging and Analysis, LemnaTec Manual, 2007.
- [Gr06] Granier, C.; et. al.: 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, *The New phytologist*, 169, 3, 2006, S.623–35.
- [Ha11] Hartmann, A.; et. al.: HTPPheno: An Image Analysis Pipeline for High-Throughput Plant Phenotyping. *BMC Bioinformatics*, 12, 2011, S.148–157.
- [KPE12] Klukas, C.; Pape, J.-M.; Entzian, A.: Analysis of high-throughput plant image data with the information system IAP, *Journal of integrative bioinformatics*, 9, 2, 2012, S.191-193.
- [Le12] Lemnatec GmbH, LemnaTec, Verfügbar: <http://www.lemnatec.com>, Zugriff: 18.10.2012
- [SP00] Sauvola, J.; Pietikäinen, M.: Adaptive document image binarization, *Pattern Recognition*, 33, 2000, S.225–236.
- [SRE12] Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W.: NIH Image to ImageJ: 25 years of image analysis, *Nature methods*, 9, 2012, S.671–675.
- [Wa07] Walter, A., et. al.: Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species, *The New phytologist*, 174, 2, 2007, S.447–455.
- [Wi09] Wickham, H.; ggplot2. Springer, New York, NY, 2009.