

Development of ground truth dataset with phenotypic data analysis and their application in agriculture and research

Desarrollo de ground truth dataset con análisis de datos fenotípicos y su aplicación en agricultura e investigación

M.V. Díaz-Galián^{1*}, P.J. Navarro Lorente², M. Egea-Cortines¹

¹Instituto de Biotecnología Vegetal, Edificio I+D+I, Universidad Politécnica de Cartagena, Campus Muralla del Mar, 30202 Cartagena, Murcia. Spain.

²Departamento de Tecnologías de la Información y las Comunicaciones, ETSIT, Universidad Politécnica de Cartagena, Campus Muralla del Mar, 30202 Cartagena, Murcia. Spain.

*mariavictoria.diaz@edu.upct.es

Abstract

Currently, new techniques to plant phenotyping, such as computer vision, have enabled to improve the detection of several parameters. Artificial intelligence needs data to train algorithms, so the creation of ground truth datasets is highly relevant to research. A dataset was developed comprising 114 flowers of *Antirrhinum majus* (commercial variety Vilmorin long variee). Furthermore, several phenotypic features were measured, such as length, width, weight, and anthocyanin content. This study confirms that weight is the best parameter to determine flower development, as well as this knowledge enables to create machine learning algorithms to an automatic and non-invasive phenotyping.

Keywords: phenotyping; computer vision; image analysis.

Resumen

Actualmente, nuevas técnicas para el fenotipado vegetal, tales como la visión artificial, han permitido mejorar la detección de características. La inteligencia artificial necesita de datos masivos para la generación de algoritmos. Por ello, la creación de ground truth datasets es altamente relevante en investigación. Un dataset fue desarrollado usando 114 flores de *Antirrhinum majus* (variedad comercial Vilmorin long variee). Además, las características fenotípicas como la longitud, peso, anchura y contenido de antocianinas fueron medidas. Se confirmó que el peso es la mejor característica para determinar el estado de desarrollo floral. Además, este conocimiento permite la creación en un futuro de programas de machine learning utilizando estos datos para un fenotipado automática y no invasivo.

Palabras clave: fenotipado; visión artificial; análisis de imagen.

1. INTRODUCTION

El fenotipo es el conjunto de todas las características externas dadas por un genotipo bajo ciertas condiciones fisiológicas, medio-ambientales y de desarrollo (1). En la actualidad, el fenotipado es capaz de detectar cambios entre mutantes y líneas silvestres, así como la calidad de las frutas y vegetales. Su identificación a través de técnicas no invasivas, como la visión por computador está siendo ampliamente extendido por los beneficios que presentan, como la no destrucción del fruto, la rapidez y la objetividad en la toma de decisión (2,3). La utilización de cámaras hiperespectrales permite adquirir imágenes con información no obtenida por otros medios como es el contenido de agua o daños en el tejido vegetal (4–8). La utilización de las imágenes obtenidas para la creación de programas de fenotipados automáticos requiere también

la obtención de datos fenotípicos con el fin de poder combinar toda la información para entrenar los algoritmos y, consecuentemente, tener los mejores resultados.

Para terminar, como el Deep learning requiere de grandes volúmenes de datos para poder funcionar de manera correcta, estos hacen uso de ground truth datasets que son conjuntos de datos con anotaciones. En nuestro caso concretamente, son imágenes asociadas a datos fenotípicos. Por esta razón, se desarrolló un dataset de imágenes y datos fenotípicos de flores *Antirrhinum*, lo cual permitirá en un futuro hacer uso de ellos y generar programas no invasivos de fenotipado.

2. MATERIALES Y MÉTODOS

2.1 Material vegetal y datos fenotípicos

Se utilizaron 114 flores de *Antirrhinum majus* en diferentes etapas de crecimiento. Se clasificaron en tres grupos: bajo (0,01-0,15 g), medio (0,15-0,3 g) y alto (0,3-0,45 g) según su peso. Se midió su longitud, anchura, peso y el contenido de antocianinas. Para la determinación de antocianinas se realizó por el método del metanol ácido como está descrito en previos estudios (9).

2.2 Adquisición de imagen

Se utilizaron dos cámaras hiperespectrales. Las imágenes fueron tomadas en una cámara diseñada para evitar los brillos y/o reflejos. Para ello, un sistema de LEDs proporcionaba las diferentes longitudes de onda necesarias para el correcto funcionamiento de las cámaras.

3. RESULTADOS Y DISCUSIÓN

En este estudio se realizó la toma de imágenes para la creación de un ground-truth dataset. Para ello se tuvieron que medir distintos parámetros fenotípicos como la longitud, anchura, peso y cantidad de antocianinas. Se observó que estas características van variando durante el desarrollo floral, por lo que pueden ser utilizadas para determinar en qué fase se encuentran. En la Figura 1 podemos observar como el peso sería la mejor característica para determinar el estado de desarrollo floral. Conocer en qué fase está es altamente importante porque permite tener una población de estudio uniforme en estudios de volátiles, por ejemplo. Los datos fenotípicos fueron analizados en detalle con diferentes líneas de tendencia (lineal, exponencial, logarítmica y polinomial) con la intención de obtener datos adicionales imposibles de adquirir con los datos crudos. La mejor línea de tendencia permitió detectar un desarrollo de doble sigmoide en el peso de las flores, como ocurre en el desarrollo de las bayas. En cuanto al contenido de antocianinas, este también iba incrementándose a lo largo del desarrollo.

Por todo lo anteriormente comentado, es interesante la creación de este dataset puesto que se podrían crear algoritmos de Deep learning para detectar la cantidad de antocianinas de una forma no invasiva y rápida. Con la información aportada, se podrían realizar programas de inteligencia artificial para determinar el desarrollo floral, lo cual es recomendable en estudios fenotípicos donde se precisa utilizar flores que se encuentren en el mismo momento del desarrollo floral. Por ello, la creación de algoritmos utilizando este ground-truth dataset permitiría de forma precisa la selección. Otras aplicaciones serían en la detección de comportamientos fenotípicos anómalos en estudios de mutantes, como previamente han sido descritos (10-12).

4. CONCLUSIONES

- Se confirma que todas las características estudiadas pueden ser correlacionadas con el desarrollo floral.
- El ground truth dataset creado puede utilizarse para generar un futuro programa de inteligencia artificial no invasivo para detectar el desarrollo floral.
- El desarrollo del peso sigue una curva doble sigmoide.

5. REFERENCIAS

1. Zhou N, Siegel ZD, Zarecor S, Lee N, Campbell DA, Andorf CM, et al. Crowdsourcing image analysis for plant phenomics to generate ground truth data for machine learning. *PLoS Comput Biol.* 2018 jul 30;14(7):e1006337.
2. Su W-H, Sun D-W. Multispectral Imaging for Plant Food Quality Analysis and Visualization. *Compr Rev Food Sci Food Saf.* 2018;17(1):220-39.
3. Xu R, Li C, Paterson AH. Multispectral imaging and unmanned aerial systems for cotton plant phenotyping. *PLoS ONE.* 27 de febrero de 2019;14(2):e0205083.
4. Danielsen S, Munk L. Evaluation of disease assessment methods in quinoa for their ability to predict yield loss caused by downy mildew. *J Crop Prot.* 2004;23(3):219-28.
5. Henebry GM. Remote sensing of vegetation: Principles, techniques, and applications. *BioScience.* 2011;61(7):568-9.
6. Navarro PJ, Miller L, Gila-Navarro A, Díaz-Galián MV, Aguila DJ, Egea-Cortines M. 3DeepM: An Ad Hoc Architecture Based on Deep Learning Methods for Multispectral Image Classification. *Remote Sensing.* 2021 jan;13(4):729.
7. Sasaki Y, Okamoto T, IMOU K, TORII T. Automatic diagnosis of plant disease recognition between healthy and diseased leaf. *Journal of the Japanese Society of Agricultural Machinery.* 1999;61(2):119-26.
8. Serrano L, González-Flor C, Gorchs G. Assessment of grape yield and composition using the reflectance based Water Index in Mediterranean rainfed vineyards. *Remote sensing of environment.* 2012;118:249-58.
9. Li Z, Zhao M, Jin J, Zhao L, Xu Z. Anthocyanins and their biosynthetic genes in three novel-colored *Rosa rugosa* cultivars and their parents. *Plant Physiology and Biochemistry.* 2018 Aug;129:421-8.
10. Bowman JL, Smyth DR, Meyerowitz EM. Genes directing flower development in *Arabidopsis*. *The Plant Cell.* 1989 jan 1;1(1):37-52.
11. Nilsson O, Weigel D. Modulating the timing of flowering. *Current Opinion in Biotechnology.* 1997 Apr 1;8(2):195-9.
12. Martin C, Prescott A, Mackay S, Bartlett J, Vrijlandt E. Control of anthocyanin biosynthesis in flowers of *Antirrhinum majus*. *Plant J.* 1991 jul;1(1):37-49.

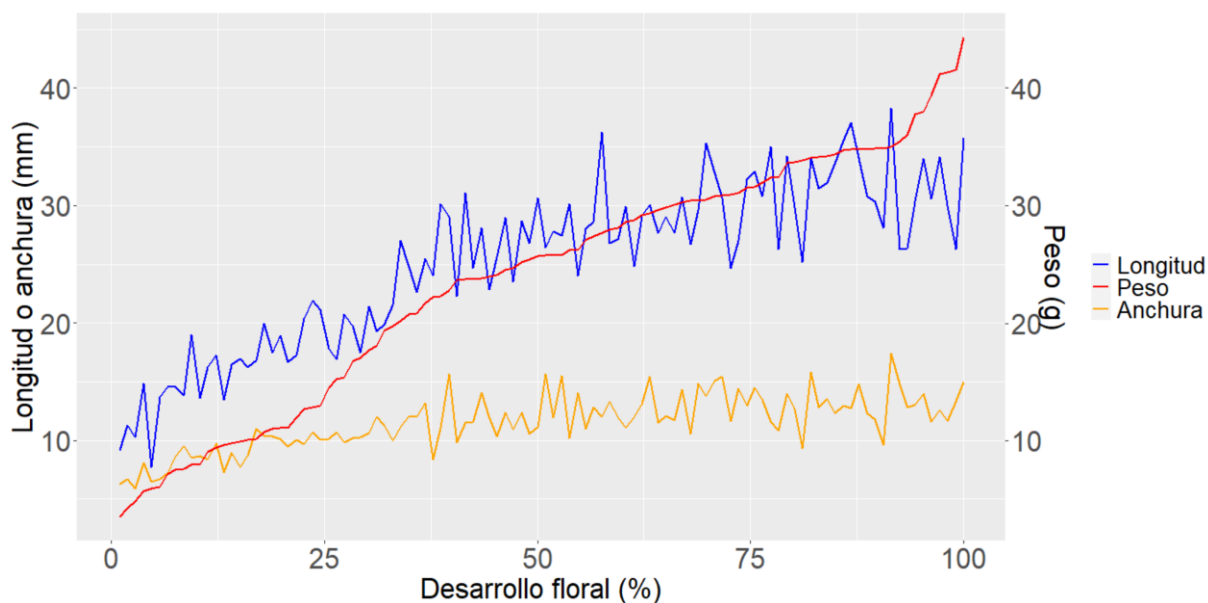


Figura 1. Datos fenotípicos del desarrollo floral de *Antirrhinum*.