

DETECCIÓN Y SEGUIMIENTO DE RACIMOS PARA FENOTIPADO *en vídeos laterales de dron*

La información sobre el fenotipado de la uva es valiosa para la toma de decisiones, ya que proporciona características del dosel vegetal, la fenología de la planta y las características foliares, que influyen en gran medida en el estado sanitario de la vid. El fenotipado mediante inspecciones sobre el terreno requiere mucho tiempo y es laborioso. Sin embargo, los avances en teledetección, como el uso de vehículos aéreos no tripulados (UAV) con múltiples tipos de sensores a bordo, ofrecen una alternativa que ahorra tiempo y trabajo. En este sentido, las técnicas de visión por ordenador, conocidas como *Computer Vision*, como la detección y el seguimiento de objetos, entran en juego como herramientas de análisis de los conjuntos de datos adquiridos con los UAV o los vehículos terrestres no tripulados.

MAR ARIZA-SENTÍS, SERGIO VÉLEZ, HILMY BAJA, JOÃO VALENTE

Information Technology Group, Wageningen University & Research, 6708 PB Wageningen, Netherlands

Estudios recientes se han centrado en el fenotipado, principalmente en nubes de puntos 3D. Rose *et al.* (2016) utilizaron un vehículo que tenía múltiples cámaras montadas en él, capturando datos 3D mediante la reconstrucción de las imágenes de las uvas utilizando nubes de puntos para obtener datos del fenotipo de la uva. Milella *et al.* (2019) utilizaron un método similar utilizando una cámara RGB-D. En su estudio se alcanzó una precisión en detección de uvas del 91% utilizando la red neuronal VGG19 (Simonyan & Zisserman, 2014).

En lo que respecta a la detección y el seguimiento de objetos, la mayoría de los marcos de *Deep Learning*, por ejemplo, Mask R-CNN (He *et al.*, 2017) y You Only Look Once (YOLO) (Redmon *et al.*, 2016), utilizan con-

juntos de datos que contienen etiquetas de personas, animales, coches y objetos comunes. Por ello, no cuentan con mucha información relativa a la agricultura, aunque algunos estudios ya han aplicado la detección de objetos en el campo de los cultivos leñosos. Stein *et al.* (2016) utilizaron Faster R-CNN en mangos, y Wang *et al.* (2019) implementaron un algoritmo basado en YOLO, también para mangos. Un estudio de Bargoti & Underwood (2017) examinó manzanas, mangos y almendras utilizando Faster R-CNN.

Por otra parte, el seguimiento y segmentación de múltiples objetos (MOTS) consiste en detectar y seguir el objeto en un vídeo, de manera que solo se detecta el objeto una vez y luego se identifica el mismo objeto a lo largo del vídeo. Voigtlaender *et al.* (2019) desarrollaron esta ta-

rea de visión por ordenador junto con el primer marco de detección y seguimiento MOTS, denominado TrackR-CNN. Ha habido estudios previos donde se ha utilizado MOTS en cultivos leñosos, como el de de Jong *et al.* (2022), donde aplicaron el algoritmo TrackR-CNN para la detección y seguimiento de manzanas y Ariza-Sentís *et al.* (2022) en viñedo. Además, *PointTrack* (Xu *et al.*, 2020) también fue implementado en este estudio mostrando resultados prometedores y potencial en la estimación del rendimiento de manzana a obtener al final de campaña. Sin embargo, también revelaron muchos retos en el uso de MOTS para el recuento y seguimiento de frutas, como la homogeneidad de las frutas, el tamaño de las frutas y el desafiante entorno del huerto.

Una técnica común utilizada en los

estudios mencionados es el uso de datos de entrada 3D de racimos para mayor precisión. Santos *et al.* (2020) realizaron un estudio sobre detección del contorno de racimos, empleando un conjunto de datos con racimos anotados llamado *Wine Grape Instance Segmentation Dataset* (WGISD), compuesto por 300 imágenes que muestran alrededor de 4000 racimos, localizados en viñedos con una producción basada en un sistema de espaldera, tomadas con dos cámaras. Por lo tanto, las imágenes tomadas son muy claras y cercanas, con una distancia de 1 metro de las uvas. Por ello, las imágenes plantean si un modelo entrenado con este conjunto de datos será lo suficientemente robusto para imágenes adquiridas desde diferentes plataformas, por ejemplo, drones. Hasta ahora, se carece de conjuntos de datos tomados desde UAV por lo que es interesante probar imágenes adquiridas con UAV, teniendo en cuenta los muchos usos que tienen (Tsouros *et al.*, 2019), y su creciente investigación y uso en la agricultura (Rejeb *et al.*, 2022).

Los aspectos más destacados de este artículo son los siguientes 1) detectar y rastrear racimos sobre videos grabados con dron en un viñedo comercial, presentando condiciones de iluminación complejas y oclusión de hojas, y 2) proporcionar rasgos de fenotipado como la longitud, anchura y forma del racimo en una etapa temprana del desarrollo del racimo. Este artículo ha sido adaptado de Ariza-Sentis *et al.* (2023).

Material y método

El experimento se llevó a cabo durante la campaña 2021, utilizando un conjunto de datos de un viñedo comercial de 1,06 ha y 8,1% de pendiente, *Vitis vinifera* cv. Loureiro, situado en 'Tomíño, Pontevedra', Galicia, España (X: 516989.02, Y: 4644806.53; ETRS89 / UTM zona 29N), propiedad de 'Bodegas Terras Gauda, S.A.', perteneciente a la denominación de origen "Rías Baixas". Los videos se grabaron utilizando

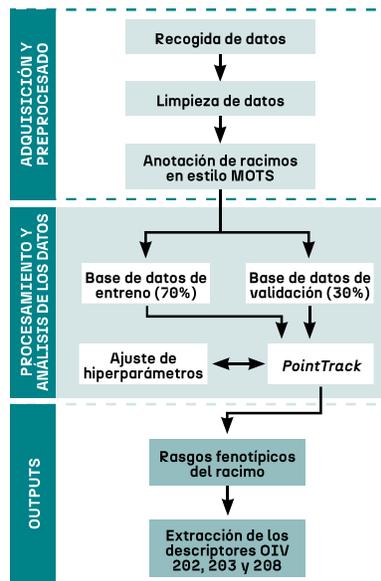


FIGURA 1

Izquierda: Flujo de trabajo de este estudio. Una vez los videos de dron han sido grabados se realiza una limpieza de los datos y anotación de los racimos presentes en los videos en estilo MOTs. El siguiente paso es el entrenamiento del algoritmo *PointTrack* y el ajuste de los hiperparámetros para la óptima detección y seguimiento de los racimos. Finalmente, los rasgos fenotípicos se extraen y se procede al conteo de los racimos presentes en cada video. Derecha: Ejemplo de video lateral grabado con el dron (arriba), ejemplo de anotaciones de racimos producidas con el software CVAT (abajo).

La OIV cuenta con numerosas normas para el ecosistema vitivinícola, una de estas normas define el fenotipo de los racimos

la plataforma UAV DJI Matrice 210 (DJI Sciences and Technologies Ltd., Shenzhen, Guangdong, China) a una velocidad de vuelo de 0,7 metros por segundo y una altitud de 3 metros sobre el nivel del mar. Los vuelos se realizaron en un día soleado, con una velocidad del viento inferior a

0,5 m/s. La cámara montada en el UAV era una DJI Zenmuse X5S. En total, se grabaron 40 videos sobre las cuatro hileras en las que se ejecutaron los vuelos. Las hileras sobre las que voló el UAV se seleccionaron en función de la fase de maduración de los racimos, para disponer de una muestra representativa del viñedo. El flujo de trabajo seguido en esta investigación se puede observar en la **Figura 1**. Los racimos se anotaron asegurándose de que sólo se anotan las uvas, sin el pedúnculo del racimo, y anotándolos si eran visibles para la cámara, incluso estando bajo una sombra. En total, se etiquetaron 29 secuencias de video para detectar y seguir los racimos, con un número total de 679 fotogramas anotados. De esos videos, se utilizó una proporción aproximada de 70/30 para el entrenamiento y la validación del algoritmo. Para la reproducibilidad y para ampliar la investigación reali-

zada en este campo, se ha publicado con acceso abierto el conjunto de datos y las etiquetas MOTs de los racimos (Ariza-Sentís *et al.*, 2023).

Fenotipado

La Organización Internacional de la Viña y el Vino (OIV) cuenta con numerosas normas para el ecosistema vitivinícola, que van desde la denominación de determinados productos hasta la clasificación de las uvas con diversos fines. Una de estas normas es un código que define el fenotipo de los racimos. Para el racimo, la longitud, anchura, y forma se definen como códigos OIV 202, 203, y 208, respectivamente.

Para evaluar los rasgos extraídos, se genera una tabla con los racimos identificados, junto con su longitud, anchura y forma (tanto en píxeles como en cm/cm²). De este modo, la conversión de píxeles a cm se realiza mediante información sobre la anchura del poste (9 cm) de cada secuencia de vídeo. Dado que los vuelos se realizaron en modo manual, cada secuencia tiene una longitud ligeramente diferente de los racimos, la relación de conversión es diferente para cada secuencia de vídeo.

De este modo, aunque los vídeos tomados desde un UAV son estables, la distancia tomada desde las filas

puede producir una pequeña variación de ángulos entre diferentes filas y, dado que los postes de las vides en las imágenes se tomaron en ángulos diferentes, existe la posibilidad de que las mediciones de los píxeles también estén sesgadas, generando ligeros errores entre la conversión de las diferentes hileras. Seethepalli *et al.* (2020) describen que si las imágenes tendrían una precisión milimétrica si tuviesen al menos 10 píxeles/milímetro, aunque este no es el caso para nuestro estudio. Por lo tanto, es posible una desviación de ~3 centímetros con respecto a las mediciones reales. Una vez se conocen las dimensiones en centímetros de la longitud y anchura del racimo, éstos se clasifican

**Esta técnica
puede utilizarse
para estimar de
forma temprana
el número de
racimos, sus
características y
su fenotipo**

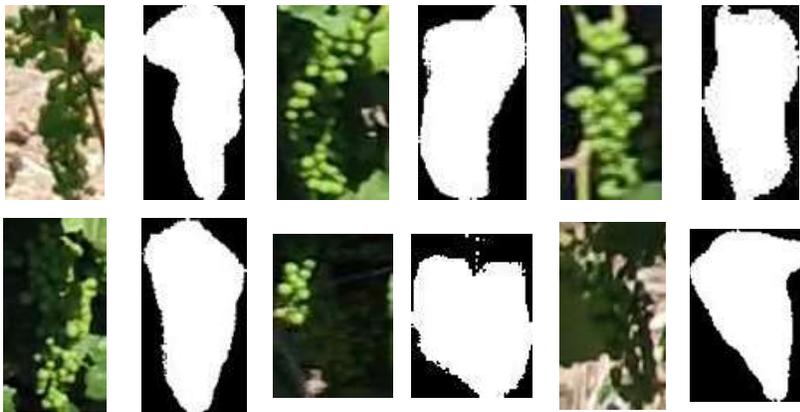


TABLA 1

Capturas de los racimos tal y como se presentan en el vídeo (columnas impares), junto con la máscara de predicción de los racimos obtenida con el algoritmo *PointTrack* (columnas pares).

Las imágenes se muestran exactamente como las genera el algoritmo, sin ningún tipo de retoque y sin alteraciones, por lo que pueden aparecer pixeladas.

en las 5 clases marcadas por la OIV (clases de 1 a 5). Referente a la forma del racimo, la OIV 208 cuenta con 3 categorías. Para poder clasificar cada racimo, se calcula la ratio entre longitud y anchura. Si se encuentra por debajo de 1,1 se clasifica como 1, mayor a 1,3 es clase 2, y finalmente entre 1,1 y 1,3, clase 3.

Resultados y discusión

El algoritmo *PointTrack* permitió la detección y seguimiento de los racimos en los vídeos, pudiéndose aislar el contorno de cada racimo. En la **Tabla 1** se puede observar cómo se ve el racimo en el vídeo, junto con la máscara de predicción obtenido por el algoritmo. Los resultados de fenotipado obtenidos con *PointTrack* se muestran en la **Tabla 2**, indicando la clasificación OIV para la longitud, anchura y forma.

Los resultados de este estudio contrastan con los de de Jong *et al.* (2022) que probaron *PointTrack* en un conjunto de datos de manzanas en un huerto obtenidos también con dron. Un factor que es bastante crucial en por qué el conjunto de datos de manzanas es más preciso se debe al color de los objetos en comparación con su entorno circundante (Bullinger *et al.*, 2017). Las manzanas tienen una tonalidad roja sobre una vegetación verde. Además, la red *PointTrack* hace hincapié en una modalidad de datos basada en la diferenciación del color de los objetos (Xu *et al.*, 2020). Esta distinción del color contribuye a la obtención de buenos resultados para la detección y seguimiento de manzana, sin embargo, las uvas son verdes sobre una vegetación con un tono similar de verde. De hecho, multitud de estudios han remarcado la importancia de las características visuales para la detección de objetos, especialmente de su contorno (García-García *et al.*, 2018; Girshick *et al.*, 2014; Zhu *et al.*, 2012). Esta falta de distinción de color entre el objeto y el entorno obstaculizó el modelo para detectar correctamente los racimos, a pesar de las numerosas estrategias diferentes aplicadas para entrenar el conjunto de datos.