

Integración de una cámara multiespectral y aprendizaje automático para clasificación de manzanas

Realizado por Bryan Josue Zambrano Vera

25 de octubre de 2023

Resumen

Este trabajo presenta un sistema capaz de realizar una clasificación binaria (buena y podrida) de manzanas red delicious, lo cual se logra mediante el uso del Índice de Vegetación de Diferencia Normalizada (NDVI) y una Red Neural Xception. Para ello, se utiliza una cámara multiespectral para observar detalles fuera del alcance del ojo humano, reduciendo la presencia de errores durante su clasificación. La selección del NDVI es resultado de su comparación con los índices de vegetación GNDVI, GVI, NDRE, NDVIR, NG, NGRDI y RVI, aplicados a un banco de imágenes obtenido mediante la cámara multiespectral. Además, se muestran los resultados de la clasificación al utilizar redes neuronales Xception, ResNet y MobileNet, lo que justifica el uso de la red Xception.

1. Introducción

Hoy en día, el almacenamiento de manzanas en cámaras de refrigeración suele presentar condiciones no favorables como el escaso mantenimiento del sistema de aire acondicionado y la recirculación de aire, lo cual permite a bacterias y hongos mantenerse y reproducirse, provocando pudrición (Paulus et al., 1997)

Por ello, surge la necesidad de construir y diseñar sistemas de clasificación para control de calidad en menor tiempo, de manera automática, con menor empleo de personal y mayor exactitud. Lo cual se puede lograr mediante la implementación de Redes Neuronales (NN por sus siglas en inglés), generando mejores estándares de calidad y ahorro económico

- Obtención y procesamiento de imágenes
- Entrenamiento y comparativa de redes neuronales
- Integración del sistema de clasificación
- Conclusiones
- Referencias

2. Obtención y procesamiento de imágenes

La presente sección, describe el proceso de obtención y procesamiento de imágenes provenientes de una CM. La cámara utilizada es la MicasenseRededge-M, la cual toma 5 capturas correspondientes a cada banda espectral: azul (B), verde (G), rojo (R), Infrarrojo Cercano (NIR) y Borde Rojo (RE). La comunicación entre la cámara y el computador se realiza vía ethernet y la programación mediante OpenCV (RedEdge, 2015).

Alineamiento y reconstrucción RGB .- La CM presenta un desfase de posición en los lentes. Por lo tanto, se procede a realizar alineamiento de las 5 bandas, teniendo como objetivo una reconstrucción centrada de la manzana.

Alineamiento .- Se realiza una transformación de perspectiva, mediante la cual se estima la orientación relativa dentro de dos imágenes. Para ello se utiliza una matriz de homografía H (Escamilla, 2012).

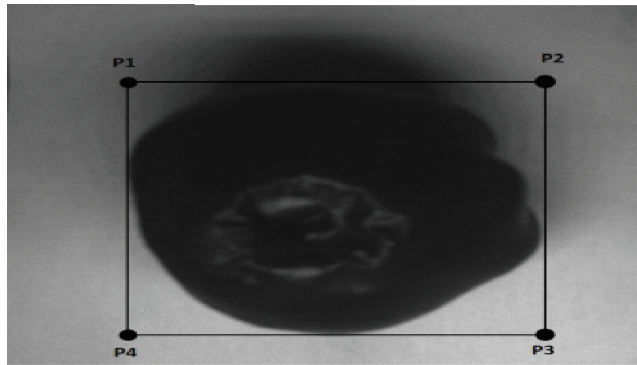
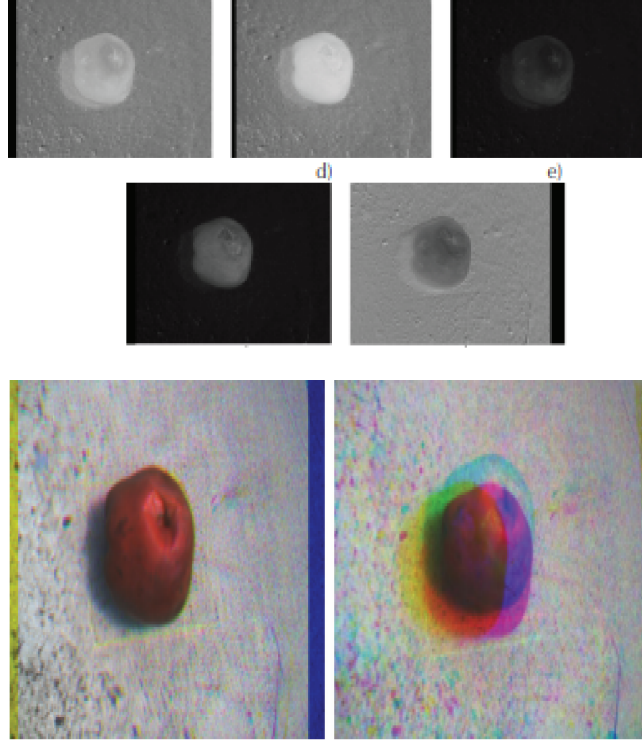


Figura 1: Selección de puntos requeridos para calcular matriz de homografía.

Tomando puntos de la banda RE como puntos de destino, se obtiene la matriz de homografía respecto a una banda distinta. Posteriormente, se hace el alineamiento de la banda seleccionada con la de RE.

Reconstrucción RGB .- Después del alineamiento se extraen los canales de la imagen. Cabe mencionar que la descomposición proporciona 3 canales. Una vez que se tienen imágenes de cada banda constituidas por un solo canal, se realiza la unión de las bandas B, G y R. Obteniéndose una imagen conformada por los valores de los 3 canales. La Figura 2 muestra el resultado del alineamiento y reconstrucción RGB.

Índice de vegetación.- Una vez concluido el alineamiento de imágenes se realizó la obtención del VI, el cual resulta de operaciones algebraicas entre 2 o



más bandas espectrales (Ali et al., 2017). Para esto, las imágenes deben estar conformadas por un canal y valores flotantes.

La Figura 3 muestra los resultados de cada VI, tanto el NDVI como el GVI destacan mejor las pudriciones (Figuras 3a y 3d). Sin embargo, el GVI presenta menor contraste de la manzana, por lo cual se optó por utilizar el NDVI.

3. Entrenamiento y comparativa de redes neuronales

Los métodos de segmentación y máscaras requieren que se dicten reglas para realizar una acción. Por el contrario, los métodos de ML requieren los datos y las respuestas para generar las reglas o algoritmos de relación entre estos (Xul et al., 2008). Las arquitecturas de NN empleadas para clasificación de imágenes (Figura 7) están compuestas por una capa de entrada, capa de salida y capas intermedias u ocultas.

Para este trabajo se empleó un conjunto de datos que contiene 816 imágenes con 425,400,3 píxeles. El banco de imágenes se dividió en dos conjuntos: entrenamiento y validación. El primer conjunto contiene 705 elementos utilizados para la etapa de entrenamiento de la NN, mientras que el segundo se confor-

ma de los 111 elementos restantes y es utilizado para la validación y pruebas en línea. Como se observa, la base de datos es pequeña, considerando trabajos como (Dargan et al., 2019) y los avances en clasificación de imágenes mediante ML. En particular, Deep Learning (ImageNet o MNIST); realiza Transferencia de Aprendizaje (en adelante TA).

Programación Red Neuronal La programación y entrenamiento de las NNs se realizó mediante la librería Keras de Python, el algoritmo utilizado se describe en la Figura 8.

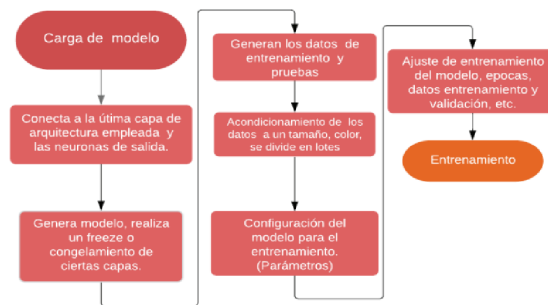


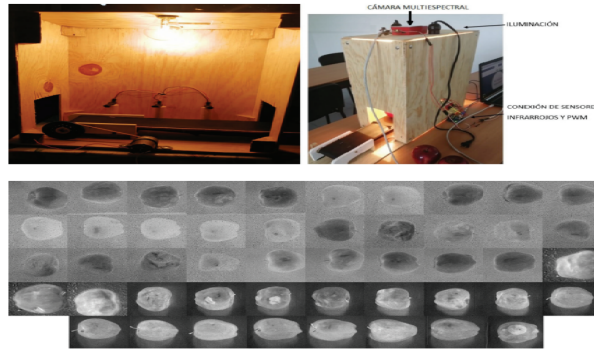
Figura 8. Etapas presentes durante el entrenamiento de las NNs

4. Integración del sistema de clasificación

Se construyó el prototipo de la Figura 12. El cual consta de una banda transportadora de 0.09 x 0.6 m, trabaja a una velocidad de 0.075 m/s y permite procesar 4 manzanas por minuto. Además, se emplea iluminación halógena en forma directa que ofrece un flujo luminoso de 1490 lúmenes. Para detectar el paso de las manzanas se emplean 3 sensores infrarrojo LM8393 acoplados a un costado de la banda, dejando una distancia de 0.09m entre cada sensor, los cuales sirven como capturador externo, es decir; cuando un sensor detecta el paso de la manzana se manda un pulso a la CM para realizar una captura

Clasificación en línea.- Se realizaron pruebas en línea para evaluar la capacidad de clasificación real del modelo propuesto, se tomó una muestra conformada por 16 manzanas o 48 imágenes.

Discusión de resultados.- El desempeño de 88.46 Singh (2019), quienes clasifican fuera de línea manzanas podridas y buenas. Singh et al. logran 98.9 utilizando una SVM y características de textura como: transformada Wavelet, histograma de gradientes orientados, energía de la textura de Law, etc. En un enfoque similar, Moallem et al. (2017) obtienen un desempeño de 92.5



5. Conclusiones

La metodología descrita reduce la necesidad de extraer características como: textura, forma, color, etcétera. En lugar de ello, se usa el NDVI para observar pudriciones. Lo anterior se logra mediante operaciones algebraicas entre las bandas espectrales proporcionadas por la CM. Por otra parte, la arquitectura Xc se implementó en línea para clasificar manzanas buenas y podridas.