



UNCUYO
UNIVERSIDAD
NACIONAL DE CUYO

Proyecto Final de Estudios

Cosechador de Lechugas Autónomo con Unidad de
Detección por Inteligencia Artificial

Brenda Gudiño
Alan Vignolo

Fecha de presentación: XX/XX/2025

Índice general

Resumen	3
1. Introducción	4
1.1. Contexto y Motivación	4
1.2. Objetivos del Proyecto	4
1.3. Alcance y Limitaciones	4
1.4. Estructura del Documento	4
2. Marco Teórico	5
2.1. Sistemas de Control Jerárquico	5
2.2. Fundamentos de Visión Artificial y Aprendizaje Profundo	5
2.2.1. Procesamiento Digital de Imágenes	5
2.2.2. Detección y Análisis de Contornos	9
2.2.3. Clasificación Basada en Análisis Morfológico	15
2.2.4. Métricas de Evaluación para Sistemas de Clasificación	21
2.3. Cinemática de Robots Cartesianos	27
2.4. Sistemas de Transmisión Mecánica	27
2.5. Control de Motores Paso a Paso	27
3. Desarrollo del Sistema	28
3.1. Arquitectura General del Sistema	28
3.2. Modelado y Diseño Mecánico	28
3.2.1. Especificaciones de Diseño y Restricciones	28
3.2.2. Análisis Cinemático del Sistema	30
3.2.3. Diseño Estructural	30
3.2.4. Sistema de Movimiento Horizontal	30
3.2.5. Sistema de Movimiento Vertical	30
3.2.6. Brazo Robótico	30
3.2.7. Modelado CAD y Fabricación	30
3.3. Sistema de Control de Bajo Nivel (Nivel Regulatorio)	30
3.3.1. Arquitectura del Nivel Regulatorio	30
3.3.2. Hardware de Control - Arduino Mega 2560	30
3.3.3. Selección y Dimensionamiento de Actuadores	30
3.3.4. Sensores de Seguridad	30
3.3.5. Control de Movimiento	30
3.3.6. Protocolo de Comunicación UART	30
3.4. Sistema de Supervisión y Alta Gestión (Nivel Supervisor)	30
3.4.1. Arquitectura del Nivel Supervisor	30
3.4.2. Hardware Supervisor - Raspberry Pi 4	30
3.4.3. Máquina de Estados Supervisora	30
3.4.4. Coordinación con Nivel Regulatorio	30
3.5. Inteligencia Artificial y Visión por Computadora	30
3.5.1. Arquitectura y Pipeline de Procesamiento	30
3.5.2. Sistema de Posicionamiento Visual	37
3.5.3. Sistema de Clasificación de Cultivos	44

Proyecto Final de Estudios: Robot cosechador automático

3.5.4. Mapeo y Optimización de Rutas	50
3.6. Interfaz de Usuario y Supervisión	56
3.7. Montaje e Integración	56
4. Pruebas y Resultados	66
4.1. Metodología de Pruebas	66
4.2. Pruebas del Sistema Mecánico	66
4.3. Pruebas del Sistema de Control	66
4.4. Pruebas del Sistema de IA	66
4.5. Pruebas de Integración	66
4.6. Análisis de Resultados	66
5. Conclusiones y Trabajo Futuro	67
5.1. Conclusiones Generales	67
5.2. Aportes del Proyecto	67
5.3. Trabajo Futuro	67
A. Diagramas Eléctricos Completos	68
B. Código Fuente Relevante	69
C. Especificaciones Técnicas de Componentes	70
D. Manual de Usuario	71
E. Hojas de Datos	72
Referencias	73

Resumen

1. Introducción

1.1. Contexto y Motivación

1.2. Objetivos del Proyecto

1.3. Alcance y Limitaciones

1.4. Estructura del Documento

2. Marco Teórico

2.1. Sistemas de Control Jerárquico

2.2. Fundamentos de Visión Artificial y Aprendizaje Profundo

2.2.1. Procesamiento Digital de Imágenes

El procesamiento digital de imágenes constituye el fundamento tecnológico de los sistemas de visión artificial aplicados a robótica agrícola. Esta disciplina permite transformar datos visuales crudos capturados por cámaras en información estructurada y procesable por algoritmos de decisión, integrando conceptos de análisis matemático, álgebra lineal y teoría de señales.

Representación Digital de Imágenes

Una imagen digital se define matemáticamente como una función bidimensional discreta $f(x, y)$, donde las coordenadas espaciales (x, y) representan posiciones en el plano de la imagen y el valor de la función indica la intensidad luminosa en cada punto. Para imágenes a color en formato RGB, esta representación se extiende a un espacio tridimensional:

$$\mathbf{I}(x, y) = \begin{bmatrix} R(x, y) \\ G(x, y) \\ B(x, y) \end{bmatrix} \in [0, 255]^3 \quad (2.1)$$

donde R , G y B representan los canales de color rojo, verde y azul respectivamente. Esta codificación permite representar aproximadamente 16.7 millones de colores distintos (256^3), cubriendo ampliamente el espectro visible necesario para aplicaciones de agricultura de precisión.

La resolución espacial de una imagen determina la cantidad de información disponible para el análisis. Una imagen de dimensiones $M \times N$ píxeles contiene $M \cdot N$ elementos de información, cada uno representando el promedio de la radiancia incidente sobre el área del sensor correspondiente a ese píxel.

Espacios de Color y Transformaciones

El espacio de color RGB, aunque intuitivo y ampliamente utilizado en sistemas de captura, presenta limitaciones para ciertas tareas de procesamiento debido a su fuerte correlación entre canales y sensibilidad a variaciones de iluminación. Por esta razón, la conversión a espacios de color alternativos resulta fundamental en visión por computadora.

Espacio HSV (Hue, Saturation, Value)

El espacio HSV separa la información cromática (matiz y saturación) de la información de iluminación (valor), proporcionando robustez frente a cambios en las condiciones de luz. La transformación de RGB a HSV se define mediante:

$$V = \max(R, G, B) \quad (2.2)$$

$$S = \begin{cases} \frac{V - \min(R, G, B)}{V} & \text{si } V \neq 0 \\ 0 & \text{si } V = 0 \end{cases} \quad (2.3)$$

$$H = \begin{cases} 60^\circ \cdot \frac{G-B}{V-\min(R,G,B)} & \text{si } V = R \\ 60^\circ \cdot \left(2 + \frac{B-R}{V-\min(R,G,B)}\right) & \text{si } V = G \\ 60^\circ \cdot \left(4 + \frac{R-G}{V-\min(R,G,B)}\right) & \text{si } V = B \end{cases} \quad (2.4)$$

Esta descomposición resulta particularmente ventajosa para la detección de elementos basada en contraste y brillo.

El Canal V como Descriptor de Intensidad

El canal V (Value) captura la intensidad luminosa independientemente del color, representando el brillo máximo entre los tres canales RGB. Esta propiedad lo convierte en un descriptor ideal para detectar objetos oscuros contra fondos claros o viceversa.

Para la detección de elementos de contraste oscuro (como marcadores negros o cintas de cultivo), el canal V permite una segmentación robusta mediante umbralización:

$$\text{Objeto oscuro} \Leftrightarrow V(x, y) < T_V \quad (2.5)$$

donde T_V es un umbral empírico que depende de las condiciones de iluminación del entorno. En aplicaciones de cultivo hidropónico bajo iluminación controlada, valores típicos son $T_V \in [30, 70]$, permitiendo detectar elementos con baja reflectancia sin ser afectados por variaciones de color.

Ventajas del Canal V para Detección

La separación de la información de brillo en el canal V ofrece ventajas operativas críticas:

- **Invariancia parcial a color:** Elementos oscuros se detectan independientemente de su matiz (negro, azul oscuro, verde oscuro)
- **Umbralización simple:** Un único parámetro T_V define la separación
- **Robustez ante variaciones uniformes:** Cambios globales de iluminación afectan proporcionalmente todos los píxeles
- **Eficiencia computacional:** Operaciones sobre un único canal en lugar de tres

Matemáticamente, el canal V satisface:

$$V(\alpha \cdot \mathbf{I}_{RGB}) = \alpha \cdot V(\mathbf{I}_{RGB}) \quad \forall \alpha > 0 \quad (2.6)$$

donde α representa un factor de escala de iluminación global. Esta linealidad permite compensar variaciones uniformes de brillo mediante ajustes del umbral.

Umbralización y Segmentación

La umbralización constituye una técnica fundamental de segmentación que particiona una imagen en regiones de interés mediante la comparación de intensidades de píxeles contra valores de referencia.

Umbralización Binaria Inversa

En escenarios donde los elementos de interés presentan intensidades menores que el fondo, la umbralización inversa resulta esencial:

$$g(x, y) = \begin{cases} 255 & \text{si } f(x, y) < T \\ 0 & \text{si } f(x, y) \geq T \end{cases} \quad (2.7)$$

Esta operación genera una imagen binaria donde los píxeles blancos (255) corresponden a regiones oscuras de la imagen original. Es la técnica empleada para detectar cintas de cultivo negras, tubos oscuros y marcadores de contraste en agricultura de precisión.

Proyecto Final de Estudios: Robot cosechador automático

Aplicación al Canal V para Detección de Oscuros

La combinación de conversión HSV y umbralización inversa del canal V proporciona un método robusto para segmentar elementos oscuros:

$$\text{Máscara}(x, y) = \begin{cases} 255 & \text{si } V(x, y) < T_V \\ 0 & \text{si } V(x, y) \geq T_V \end{cases} \quad (2.8)$$

Por ejemplo, con $T_V = 50$, se destacan todos los píxeles con valor de brillo inferior a 50, típicamente correspondientes a superficies oscuras como:

- Cintas de identificación negras
- Tubos de cultivo sin contenido
- Tierra o sustrato oscuro
- Sombras pronunciadas

Selección Empírica del Umbral

A diferencia de métodos adaptativos como Otsu, en entornos controlados resulta más efectivo establecer umbrales fijos determinados experimentalmente. El procedimiento consiste en:

1. Capturar imágenes representativas del escenario operativo
2. Analizar histograma del canal V para objetos de interés
3. Identificar el rango de valores V que caracteriza consistentemente dichos objetos
4. Establecer T_V en el punto que maximiza separación entre objeto y fondo
5. Validar con conjunto de prueba diverso

Este enfoque aprovecha el conocimiento específico del entorno operativo, resultando en umbrales más robustos que métodos automáticos que no consideran la naturaleza del problema.

Operaciones Morfológicas

Las operaciones morfológicas manipulan la forma y estructura de objetos en imágenes binarias mediante la aplicación de elementos estructurantes. Se fundamentan en la teoría de conjuntos y permiten eliminar ruido, rellenar huecos y separar objetos conectados.

Erosión

La erosión reduce el tamaño de los objetos eliminando píxeles del perímetro:

$$(A \ominus B)(x, y) = \min_{(s,t) \in B} \{A(x + s, y + t)\} \quad (2.9)$$

donde A es la imagen binaria y B es el elemento estructurante. Esta operación elimina pequeñas protuberancias y separa objetos débilmente conectados.

Dilatación

La dilatación expande los objetos añadiendo píxeles al perímetro:

$$(A \oplus B)(x, y) = \max_{(s,t) \in B} \{A(x + s, y + t)\} \quad (2.10)$$

Esta operación rellena pequeños huecos y conecta regiones próximas.

Apertura y Cierre

Proyecto Final de Estudios: Robot cosechador automático

Estas operaciones compuestas combinan erosión y dilatación para refinar máscaras:

$$\text{Apertura: } A \circ B = (A \ominus B) \oplus B \quad (2.11)$$

$$\text{Cierre: } A \bullet B = (A \oplus B) \ominus B \quad (2.12)$$

La apertura elimina pequeños objetos y protuberancias manteniendo el tamaño aproximado de objetos grandes. El cierre rellena huecos pequeños y conecta componentes próximos, resultando esencial para consolidar regiones fragmentadas.

Elementos Estructurantes

La forma y tamaño del elemento estructurante B determina el comportamiento de las operaciones morfológicas:

- **Cuadrado** 3×3 : Operaciones isotrópicas básicas
- **Cuadrado** 5×5 : Mayor alcance, elimina ruido moderado
- **Rectangular vertical** 7×3 : Conectividad preferencial vertical, útil para unir fragmentos de objetos alargados verticalmente

La secuencia típica de refinamiento morfológico en aplicaciones de cultivo es:

1. Apertura (3×3 , 2 iteraciones) - elimina ruido puntual
2. Cierre (3×3 , 2 iteraciones) - rellena huecos pequeños
3. Cierre (5×5 , 2 iteraciones) - conecta regiones próximas
4. Cierre vertical (7×3 , 1 iteración) - unifica objetos verticalmente fragmentados

Calibración y Transformación de Coordenadas

La calibración establece la correspondencia entre coordenadas píxel en la imagen y coordenadas físicas en el espacio de trabajo. Esta transformación es fundamental para convertir mediciones visuales en comandos de movimiento precisos.

Modelo de Transformación Lineal

Para configuraciones donde la cámara observa perpendicular al plano de trabajo a distancia constante, la transformación píxel-a-métrica puede aproximarse mediante un modelo lineal afín:

$$d_{mm} = a \cdot d_{px} + b \quad (2.13)$$

donde d_{px} es la distancia medida en píxeles, d_{mm} la distancia física en milímetros, y a, b son parámetros de calibración. El coeficiente a representa la escala (mm/píxel) y b un offset sistemático.

Para sistemas bidimensionales, la transformación se expresa como:

$$\begin{bmatrix} x_{mm} \\ y_{mm} \end{bmatrix} = \begin{bmatrix} a_x & 0 \\ 0 & a_y \end{bmatrix} \begin{bmatrix} x_{px} \\ y_{px} \end{bmatrix} + \begin{bmatrix} b_x \\ b_y \end{bmatrix} \quad (2.14)$$

donde a_x, a_y son factores de escala independientes (permitiendo píxeles no cuadrados o perspectiva leve) y b_x, b_y son desplazamientos de origen.

Estimación de Parámetros por Mínimos Cuadrados

Los parámetros de calibración se estiman mediante regresión lineal a partir de pares de puntos de correspondencia conocidos. Dado un conjunto de N mediciones:

$$\{(d_{px,i}, d_{mm,i})\}_{i=1}^N \quad (2.15)$$

el problema de mínimos cuadrados minimiza el error cuadrático:

$$\min_{a,b} \sum_{i=1}^N (d_{mm,i} - (a \cdot d_{px,i} + b))^2 \quad (2.16)$$

La solución analítica se obtiene mediante las ecuaciones normales:

$$a = \frac{N \sum d_{px,i} d_{mm,i} - \sum d_{px,i} \sum d_{mm,i}}{N \sum d_{px,i}^2 - (\sum d_{px,i})^2} \quad (2.17)$$

$$b = \frac{\sum d_{mm,i} - a \sum d_{px,i}}{N} \quad (2.18)$$

Este método garantiza la solución óptima en el sentido de mínimos cuadrados, minimizando el error cuadrático medio.

Procedimiento de Calibración

El proceso de calibración consiste en:

1. Colocar marcadores de referencia en el espacio de trabajo en posiciones conocidas
2. Capturar imagen y detectar posiciones de marcadores en píxeles
3. Medir físicamente distancias entre marcadores (con calibre o regla)
4. Calcular distancias en píxeles entre los mismos marcadores en la imagen
5. Aplicar regresión lineal para estimar parámetros a, b
6. Validar con mediciones independientes

Validación Estadística

La precisión de la calibración se cuantifica mediante el error cuadrático medio (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_{mm,i} - \hat{d}_{mm,i})^2} \quad (2.19)$$

donde $\hat{d}_{mm,i} = a \cdot d_{px,i} + b$ son las predicciones del modelo. Valores de RMSE inferiores a 1-2mm resultan aceptables para aplicaciones de manipulación agrícola de precisión, considerando las tolerancias biológicas de los cultivos.

El coeficiente de determinación R^2 complementa la evaluación:

$$R^2 = 1 - \frac{\sum (d_{mm,i} - \hat{d}_{mm,i})^2}{\sum (d_{mm,i} - \bar{d}_{mm})^2} \quad (2.20)$$

donde \bar{d}_{mm} es la media de las distancias reales. Valores $R^2 > 0,99$ indican excelente ajuste lineal, validando la aproximación del modelo.

2.2.2. Detección y Análisis de Contornos

La detección de contornos constituye una técnica fundamental en visión por computadora para la identificación y caracterización de objetos en imágenes. Los contornos representan las fronteras entre regiones con propiedades visuales distintas, proporcionando información estructural esencial para el reconocimiento y clasificación de elementos.

Definición Matemática de Contorno

Un contorno se define como la secuencia ordenada de píxeles que forman el límite de una región conexa en una imagen binaria:

$$C = \{(x_i, y_i)\}_{i=1}^n \quad \text{donde } (x_i, y_i) \in \partial R \quad (2.21)$$

siendo ∂R la frontera topológica de la región R . Formalmente, un punto (x, y) pertenece a ∂R si es parte de la región pero tiene al menos un vecino que no pertenece a ella:

$$(x, y) \in R \quad \wedge \quad \exists (x', y') \in \mathcal{N}(x, y) : (x', y') \notin R \quad (2.22)$$

donde $\mathcal{N}(x, y)$ representa el vecindario del píxel. La definición de vecindario determina la conectividad del contorno:

- **4-conectividad:** Vecinos arriba, abajo, izquierda, derecha
- **8-conectividad:** Incluye diagonales, genera contornos más suaves

Algoritmo de Detección de Contornos de Suzuki-Abe

El algoritmo de Suzuki-Abe detecta simultáneamente contornos externos e internos (huecos), organizándolos jerárquicamente. Es el método implementado en OpenCV mediante la función `findContours`.

Principio de Funcionamiento

El algoritmo recorre la imagen binaria en orden raster (izquierda-derecha, arriba-abajo) y, al encontrar un píxel de objeto aún no visitado, inicia el rastreo del contorno. Define dos tipos de bordes:

- **Contorno externo:** Límite exterior de una región de objeto
- **Contorno interno:** Límite de un hueco dentro del objeto

La jerarquía se representa mediante relaciones padre-hijo:

$$\text{Jerarquía}_i = \{\text{Siguiente}, \text{Previo}, \text{Primer_Hijo}, \text{Padre}\} \quad (2.23)$$

donde cada elemento apunta al índice del contorno correspondiente en la lista.

Modos de Recuperación

Los modos de recuperación determinan qué contornos se extraen:

- **RETR_EXTERNAL:** Solo contornos externos de nivel superior
- **RETR_LIST:** Todos los contornos sin jerarquía
- **RETR_TREE:** Jerarquía completa de contornos anidados

Para aplicaciones de cultivo donde interesa detectar objetos individuales sin relaciones jerárquicas complejas, **RETR_EXTERNAL** resulta suficiente y eficiente.

Métodos de Aproximación

El parámetro de aproximación controla la representación del contorno:

- **CHAIN_APPROX_NONE:** Almacena todos los puntos del contorno
- **CHAIN_APPROX_SIMPLE:** Comprime segmentos horizontales/verticales

CHAIN_APPROX_NONE preserva toda la información geométrica, necesario para análisis morfológico detallado.

Descriptores Geométricos de Contornos

Los descriptores cuantifican propiedades estructurales de los contornos, permitiendo clasificación y reconocimiento de formas.

Área

El área encerrada por un contorno se calcula mediante la fórmula del Shoelace (algoritmo de Gauss):

$$A = \frac{1}{2} \left| \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i) \right| \quad (2.24)$$

donde $(x_n, y_n) = (x_0, y_0)$ para cerrar el contorno. Esta fórmula es eficiente computacionalmente ($\mathcal{O}(n)$) y proporciona el área exacta del polígono.

Perímetro

La longitud del contorno se calcula como:

$$P = \sum_{i=0}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad (2.25)$$

El perímetro es sensible a la resolución de la imagen y al ruido del contorno.

Rectángulo Delimitador (Bounding Box)

El bounding box es el rectángulo alineado a ejes que contiene completamente al contorno:

$$\text{BBox} = (x_{\min}, y_{\min}, w, h) \quad (2.26)$$

donde:

$$x_{\min} = \min_i x_i, \quad y_{\min} = \min_i y_i \quad (2.27)$$

$$w = \max_i x_i - x_{\min}, \quad h = \max_i y_i - y_{\min} \quad (2.28)$$

Centroide

El centro de masa del contorno se calcula como:

$$\bar{x} = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad (2.29)$$

$$\bar{y} = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad (2.30)$$

donde A es el área calculada previamente.

Filtrado de Contornos

En aplicaciones reales, las imágenes binarias contienen múltiples contornos, muchos correspondientes a ruido o elementos irrelevantes. El filtrado elimina contornos no deseados según criterios específicos del problema.

Filtrado por Área

Eliminar contornos cuya área esté fuera de un rango esperado:

$$\text{Contorno válido} \Leftrightarrow A_{\min} \leq A_{\text{contorno}} \leq A_{\max} \quad (2.31)$$

Los umbrales se establecen según el tamaño esperado de objetos de interés. Por ejemplo, para detectar plantas en tubos de cultivo de área conocida, se pueden establecer límites adaptativos:

$$A_{min} = 0,01 \cdot A_{imagen}, \quad A_{max} = 0,6 \cdot A_{imagen} \quad (2.32)$$

filtrando tanto ruido puntual como regiones que abarcan casi toda la imagen.

Filtrado por Posición

Descartar contornos que tocan los bordes de la imagen, ya que generalmente están incompletos:

$$\text{Toca borde} \Leftrightarrow \exists (x_i, y_i) \in C : x_i < \delta \vee x_i > W - \delta \vee y_i < \delta \vee y_i > H - \delta \quad (2.33)$$

donde W, H son dimensiones de la imagen y δ es un margen de tolerancia (típicamente 5 píxeles). Adicionalmente, se puede verificar proximidad a un margen interno:

$$\text{Cerca de borde} \Leftrightarrow x_{min} < m_W \vee x_{max} > W - m_W \vee y_{min} < m_H \vee y_{max} > H - m_H \quad (2.34)$$

donde $m_W = 0,12W$ y $m_H = 0,12H$ definen un margen del 12 % desde cada borde.

Filtrado por Inclusión del Centro

Priorizar contornos que encierran el punto central de la imagen, asumiendo que el objeto de interés está centrado:

$$\text{Encierra centro} \Leftrightarrow (W/2, H/2) \in \text{Interior}(C) \quad (2.35)$$

Esto se verifica mediante el algoritmo de ray casting o la función `pointPolygonTest` de OpenCV.

Análisis de Regiones Parciales del Contorno

En muchas aplicaciones, el análisis de regiones específicas del contorno proporciona información más robusta que el contorno completo. Esto es particularmente útil cuando los objetos tienen geometría variable en diferentes zonas.

Extracción del Segmento Inferior

Para objetos verticales donde la base tiene geometría más consistente (como cintas de cultivo en tubos), analizar solo el 10 % inferior mejora la robustez:

$$\text{Región base} = \{(x, y) \in C : y \geq y_{max} - 0,1 \cdot h\} \quad (2.36)$$

donde $h = y_{max} - y_{min}$ es la altura del bounding box.

Ancho Efectivo de la Base

El ancho de la base se calcula como el rango horizontal de píxeles en la región inferior:

$$w_{base} = \max_{(x,y) \in \text{base}} x - \min_{(x,y) \in \text{base}} x + 1 \quad (2.37)$$

Este descriptor es más estable que el ancho del bounding box completo, especialmente cuando la parte superior del objeto es irregular.

Consistencia de Ancho

La varianza del ancho a lo largo de la región base cuantifica la regularidad geométrica:

$$\text{Consistencia} = 1 - \frac{\sigma_w}{\bar{w}} \quad (2.38)$$

donde σ_w es la desviación estándar de anchos por fila y \bar{w} el ancho promedio. Valores cercanos a 1 indican bases rectangulares consistentes.

Ocupación de la Base

La proporción del rectángulo base ocupada por píxeles del contorno:

$$\text{Ocupación} = \frac{N_{\text{píxeles_base}}}{w_{\text{base}} \times h_{\text{base}}} \quad (2.39)$$

Bases sólidas tienen ocupación ≥ 0.8 , mientras formas irregulares tienen menor ocupación.

Métricas de Forma para Clasificación

Rectangularidad de la Base

Mide qué tan rectangular es la región inferior del contorno:

$$\text{Rectangularidad} = \frac{A_{\text{contorno_base}}}{A_{\text{bbox_base}}} \quad (2.40)$$

donde $A_{\text{contorno_base}}$ es el área del contorno en la región base y $A_{\text{bbox_base}}$ el área del rectángulo delimitador de esa región. Valores ≥ 0.9 indican bases casi rectangulares, típicas de cintas o marcadores fabricados.

Rectitud de la Base

Evalúa qué tan horizontal es el borde inferior:

$$\text{Rectitud} = 1 - \frac{\sigma_y}{h_{\text{tolerancia}}} \quad (2.41)$$

donde σ_y es la desviación estándar de las coordenadas y de los píxeles de la base, y $h_{\text{tolerancia}}$ es un valor de referencia (ej. 5 píxeles). Bases perfectamente horizontales tienen $\sigma_y \approx 0$.

Análisis Estadístico de Píxeles Internos

Además de la geometría del contorno, el contenido interno proporciona información cromática para clasificación.

Creación de Máscara de Región

Dada una imagen multicanal \mathbf{I} y un contorno C , se crea una máscara binaria:

$$M(x, y) = \begin{cases} 255 & \text{si } (x, y) \in \text{Interior}(C) \\ 0 & \text{en caso contrario} \end{cases} \quad (2.42)$$

Ratios de Color

Para clasificar entre objetos con diferente composición cromática (ej. lechugas verdes vs. tubos negros), se calculan proporciones de píxeles de cada clase dentro del contorno.

Dadas máscaras de color verde M_G y negro M_B :

$$\text{Ratio verde} = \frac{\sum_{(x,y)} M(x, y) \wedge M_G(x, y)}{\sum_{(x,y)} M(x, y)} \quad (2.43)$$

$$\text{Ratio negro} = \frac{\sum_{(x,y)} M(x, y) \wedge M_B(x, y)}{\sum_{(x,y)} M(x, y)} \quad (2.44)$$

donde \wedge denota operación AND bit a bit.

Intensidad Promedio

El brillo medio de píxeles internos:

$$\bar{I} = \frac{\sum_{(x,y): M(x,y) > 0} I_{\text{gray}}(x, y)}{\sum_{(x,y)} M(x, y)} \quad (2.45)$$

Esta métrica complementa los ratios de color para mejorar la discriminación.

Proyecto Final de Estudios: Robot cosechador automático

Score Combinado para Selección

Cuando múltiples contornos cumplen filtros básicos, se requiere un criterio cuantitativo para seleccionar el más relevante. Un score combinado pondera múltiples descriptores:

$$S_{total} = w_1 S_{ancho} + w_2 S_{consistencia} + w_3 S_{ocupación} + w_4 S_{rectitud} + S_{bonus} \quad (2.46)$$

donde cada $S_i \in [0, 1]$ es un score normalizado y w_i son pesos que reflejan importancia relativa.

Normalización de Scores

Los descriptores se normalizan a $[0, 1]$:

$$S_{ancho} = \min \left(\frac{w_{base}}{w_{ref}}, 1 \right) \quad (2.47)$$

$$S_{consistencia} = \max \left(0, 1 - \frac{\sigma_w}{100} \right) \quad (2.48)$$

$$S_{ocupación} = \min \left(\frac{N_{píxeles}}{w_{base} \cdot h_{base}}, 1 \right) \quad (2.49)$$

Ejemplo de Ponderación

Para detección de cintas de cultivo:

$$S_{total} = 0,40 \cdot S_{ancho} + 0,30 \cdot S_{consistencia} + 0,20 \cdot S_{ocupación} + 0,10 \cdot S_{rectitud} + S_{bonus} \quad (2.50)$$

El ancho base tiene mayor peso (40 %) por ser el descriptor más discriminativo. El bonus adicional puede otorgarse por criterios secundarios (ej. estar centrado en la imagen).

Componentes Conectados

Antes de la detección de contornos, identificar componentes conectados permite filtrar regiones por propiedades antes de extraer contornos detallados.

Algoritmo de Etiquetado

El algoritmo de componentes conectados asigna una etiqueta única a cada región conexa:

$$L : \mathbb{Z}^2 \rightarrow \mathbb{N} \quad (2.51)$$

donde $L(x, y)$ es la etiqueta del píxel en (x, y) .

Estadísticas de Componentes

Para cada componente i , se extraen:

- Área: $A_i = \sum_{(x,y)} \mathbb{1}[L(x, y) = i]$
- Bounding box: (x_i, y_i, w_i, h_i)
- Centroide: (\bar{x}_i, \bar{y}_i)

Filtrado Pre-Contorno

Eliminar componentes por área antes de extraer contornos reduce carga computacional:

$$\text{Imagen filtrada}(x, y) = \begin{cases} 255 & \text{si } L(x, y) = i \wedge A_{min} \leq A_i \leq A_{max} \\ 0 & \text{en caso contrario} \end{cases} \quad (2.52)$$

Este pre-filtrado es especialmente útil cuando hay cientos de componentes pequeños de ruido.

2.2.3. Clasificación Basada en Análisis Morfológico

La clasificación morfológica de objetos utiliza características geométricas y estadísticas extraídas de contornos e imágenes para asignar categorías a elementos detectados. A diferencia de métodos de aprendizaje profundo, este enfoque se fundamenta en reglas explícitas derivadas del conocimiento del dominio y análisis estadístico de datos representativos.

Fundamentos de Clasificación por Umbrales

El método más directo de clasificación morfológica establece límites de decisión basados en valores de descriptores geométricos. Para un descriptor d y un conjunto de clases $\mathcal{C} = \{c_1, c_2, \dots, c_n\}$, la regla de clasificación se define como:

$$\text{Clase}(x) = c_i \quad \text{si} \quad t_{i-1} < d(x) \leq t_i \quad (2.53)$$

donde t_0, t_1, \dots, t_n son umbrales que particionan el espacio de características.

Clasificación por Área

Para objetos que se diferencian principalmente por tamaño, el área en píxeles constituye un descriptor efectivo:

$$\text{Clase}(C) = \begin{cases} \text{Pequeño} & \text{si } A < t_1 \\ \text{Mediano} & \text{si } t_1 \leq A < t_2 \\ \text{Grande} & \text{si } A \geq t_2 \end{cases} \quad (2.54)$$

La robustez del método depende de la separabilidad entre clases. La distancia entre centroides de clases adyacentes debe ser significativa respecto a la dispersión intra-clase.

Análisis Estadístico para Determinación de Umbrales

La selección óptima de umbrales requiere análisis estadístico de una base de datos representativa de cada clase.

Parámetros Estadísticos por Clase

Para cada clase c_i , se calculan la media y desviación estándar del descriptor de interés:

$$\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} d_j \quad (2.55)$$

$$\sigma_i = \sqrt{\frac{1}{N_i - 1} \sum_{j=1}^{N_i} (d_j - \mu_i)^2} \quad (2.56)$$

donde N_i es el número de muestras de la clase i , μ_i es la media y σ_i la desviación estándar del descriptor d .

Determinación de Umbrales mediante Análisis de Distribuciones

Asumiendo distribuciones normales para cada clase:

$$P(d|c_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(d - \mu_i)^2}{2\sigma_i^2}\right) \quad (2.57)$$

Para distribuciones con igual varianza ($\sigma_i = \sigma_{i+1} = \sigma$), el umbral óptimo entre clases consecutivas es:

$$t_i^* = \frac{\mu_i + \mu_{i+1}}{2} \quad (2.58)$$

Este umbral minimiza la probabilidad de error cuando las clases tienen igual probabilidad a priori y varianzas similares.

Clasificación Multi-Criterio con Ratios de Color

Cuando un único descriptor geométrico no proporciona separabilidad suficiente, se emplean múltiples características simultáneamente. En el contexto de agricultura de precisión, los ratios de color resultan particularmente efectivos.

Espacio de Características de Color

Dado un contorno, se calculan ratios que cuantifican la proporción de diferentes componentes cromáticos:

$$r_{verde} = \frac{N_{píxeles_verdes}}{N_{píxeles_totales}} \quad (2.59)$$

$$r_{negro} = \frac{N_{píxeles_negros}}{N_{píxeles_totales}} \quad (2.60)$$

donde $N_{píxeles_totales} = N_{píxeles_verdes} + N_{píxeles_negros} + N_{píxeles_otros}$.

Cada objeto se representa como un vector en el espacio de características:

$$\mathbf{x} = [r_{verde}, r_{negro}, I_{media}]^T \quad (2.61)$$

donde I_{media} es la intensidad promedio de píxeles internos.

Prototipos de Clase

Para cada clase c_i , se define un prototipo mediante las medias de los descriptores:

$$\boldsymbol{\mu}_i = [\mu_{verde,i}, \mu_{negro,i}, \mu_{intensidad,i}]^T \quad (2.62)$$

Desviaciones Estándar por Clase

Las variaciones dentro de cada clase se cuantifican mediante:

$$\boldsymbol{\sigma}_i = [\sigma_{verde,i}, \sigma_{negro,i}, \sigma_{intensidad,i}]^T \quad (2.63)$$

Distancia Euclidiana Normalizada

Para clasificar un objeto nuevo, se calcula su distancia a cada prototipo de clase. La normalización por desviación estándar garantiza que descriptores con diferentes rangos contribuyan equitativamente.

Distancia Normalizada

La distancia de un objeto \mathbf{x} al prototipo de la clase c_i se define como:

$$D_i(\mathbf{x}) = \sqrt{\sum_{k=1}^m \left(\frac{x_k - \mu_{k,i}}{\sigma_{k,i} + \epsilon} \right)^2} \quad (2.64)$$

donde:

- m es el número de descriptores (ej. 3 para verde, negro, intensidad)
- x_k es el valor del descriptor k para el objeto
- $\mu_{k,i}$ es la media del descriptor k para la clase i
- $\sigma_{k,i}$ es la desviación estándar del descriptor k para la clase i

Proyecto Final de Estudios: Robot cosechador automático

- ϵ es un término pequeño (ej. 10^{-6}) para evitar división por cero

Regla de Clasificación por Mínima Distancia

El objeto se asigna a la clase con menor distancia normalizada:

$$\text{Clase}(\mathbf{x}) = \arg \min_i D_i(\mathbf{x}) \quad (2.65)$$

Este criterio implementa un clasificador de vecino más cercano en el espacio normalizado.

Cálculo de Confianza

La confianza de la clasificación se deriva de las distancias relativas:

$$\text{Confianza} = 1 - \frac{D_{\min}}{D_{\max} + D_{\min}} \quad (2.66)$$

donde D_{\min} es la distancia a la clase predicha y D_{\max} la distancia a la clase más lejana. Valores cercanos a 1 indican alta confianza; valores cercanos a 0.5 sugieren ambigüedad.

Para mejorar robustez, la confianza se limita:

$$\text{Confianza}_{\text{final}} = \max(0,3, \min(1,0, \text{Confianza})) \quad (2.67)$$

garantizando un rango $[0,3, 1,0]$.

Ejemplo: Clasificación de Cultivos

Para un sistema de clasificación de tubos de cultivo con tres clases:

- **LECHUGAS**: Alto ratio verde ($r_{\text{verde}} > 0,40$)
- **PLANTINES**: Ratio verde moderado, alto ratio negro
- **VASOS** (vacíos): Muy bajo ratio verde, alto ratio negro

Prototipos Estimados

A partir de un dataset de entrenamiento:

$$\mu_{\text{LECHUGAS}} = [0,65, 0,35, 120]^T, \quad \sigma_{\text{LECHUGAS}} = [0,15, 0,15, 30]^T \quad (2.68)$$

$$\mu_{\text{PLANTINES}} = [0,45, 0,55, 90]^T, \quad \sigma_{\text{PLANTINES}} = [0,15, 0,15, 25]^T \quad (2.69)$$

$$\mu_{\text{VASOS}} = [0,05, 0,95, 50]^T, \quad \sigma_{\text{VASOS}} = [0,10, 0,10, 20]^T \quad (2.70)$$

Clasificación de Muestra

Dado un objeto con descriptores $\mathbf{x} = [0,60, 0,40, 115]^T$:

$$D_{\text{LECHUGAS}} = \sqrt{\left(\frac{0,60 - 0,65}{0,15}\right)^2 + \left(\frac{0,40 - 0,35}{0,15}\right)^2 + \left(\frac{115 - 120}{30}\right)^2} = 0,47 \quad (2.71)$$

$$D_{\text{PLANTINES}} = \sqrt{\left(\frac{0,60 - 0,45}{0,15}\right)^2 + \left(\frac{0,40 - 0,55}{0,15}\right)^2 + \left(\frac{115 - 90}{25}\right)^2} = 1,52 \quad (2.72)$$

$$D_{\text{VASOS}} = \sqrt{\left(\frac{0,60 - 0,05}{0,10}\right)^2 + \left(\frac{0,40 - 0,95}{0,10}\right)^2 + \left(\frac{115 - 50}{20}\right)^2} = 8,56 \quad (2.73)$$

La clase predicha es LECHUGAS ($D_{\min} = 0,47$) con confianza:

$$\text{Confianza} = 1 - \frac{0,47}{8,56 + 0,47} = 0,95 \quad (2.74)$$

Análisis de Separabilidad de Clases

La efectividad de un conjunto de descriptores para clasificación se cuantifica mediante métricas de separabilidad.

Criterio de Fisher

Mide la razón entre varianza inter-clase e intra-clase para un descriptor d :

$$J_F = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (2.75)$$

Valores altos de J_F indican buena separabilidad. Para múltiples clases:

$$J_F = \frac{\sum_{i=1}^C N_i (\mu_i - \mu_{global})^2}{\sum_{i=1}^C N_i \sigma_i^2} \quad (2.76)$$

donde μ_{global} es la media global ponderada.

Índice de Solapamiento

Para distribuciones gaussianas, el solapamiento entre clases i y j se aproxima:

$$\text{Overlap}_{i,j} \approx 2\Phi \left(-\frac{|\mu_i - \mu_j|}{2\sqrt{\sigma_i^2 + \sigma_j^2}} \right) \quad (2.77)$$

donde Φ es la función de distribución acumulada normal estándar. Valores cercanos a 0 indican separación perfecta; valores cercanos a 1 indican fuerte solapamiento.

Validación Estadística del Clasificador

Exactitud Estimada

La exactitud de un clasificador se define como:

$$\text{Exactitud} = \frac{N_{correctas}}{N_{total}} \quad (2.78)$$

donde $N_{correctas}$ es el número de clasificaciones correctas y N_{total} el tamaño del conjunto de prueba.

Intervalo de Confianza

El intervalo de confianza al 95 % para la exactitud es:

$$\text{IC}_{95\%} = \hat{p} \pm 1,96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \quad (2.79)$$

donde \hat{p} es la exactitud observada y n el tamaño de la muestra de prueba.

Tamaño de Muestra Requerido

Para estimar la exactitud con error máximo E y nivel de confianza del 95 %:

$$n_{min} = \left(\frac{1,96 \cdot \sqrt{\hat{p}(1 - \hat{p})}}{E} \right)^2 \quad (2.80)$$

Por ejemplo, para $E = 0,05$ (error de 5 %) y $\hat{p} = 0,9$:

$$n_{min} = \left(\frac{1,96 \cdot \sqrt{0,9 \cdot 0,1}}{0,05} \right)^2 \approx 139 \quad (2.81)$$

Se requieren al menos 139 muestras de prueba.

Robustez y Generalización

Validación Cruzada k-fold

Para evaluar la generalización del clasificador, se emplea validación cruzada:

1. Dividir dataset en k particiones de igual tamaño (típicamente $k = 5$ o $k = 10$)
2. Para $i = 1$ a k :
 - Usar partición i como conjunto de prueba
 - Usar restantes $k - 1$ particiones para calcular prototipos (μ_i, σ_i)
 - Evaluar exactitud en partición i
3. Calcular exactitud promedio

La exactitud estimada por validación cruzada es:

$$\text{Exactitud}_{CV} = \frac{1}{k} \sum_{i=1}^k \text{Exactitud}_i \quad (2.82)$$

Esta métrica proporciona una estimación más realista del desempeño en datos nuevos que la evaluación en un único conjunto de prueba.

Sensibilidad a Variaciones

La robustez ante ruido se evalúa añadiendo perturbaciones controladas a los descriptores:

$$d_{ruidoso} = d_{verdadero} + \mathcal{N}(0, \sigma_{ruido}^2) \quad (2.83)$$

donde $\mathcal{N}(0, \sigma_{ruido}^2)$ es ruido gaussiano con desviación estándar σ_{ruido} .

Un clasificador robusto mantiene exactitud alta incluso con σ_{ruido} moderado. La degradación de desempeño se cuantifica como:

$$\Delta \text{Exactitud} = \text{Exactitud}_{limpio} - \text{Exactitud}_{ruidoso} \quad (2.84)$$

Degradaciones $\Delta \text{Exactitud} < 0,1$ (10 %) indican robustez aceptable.

Ventajas y Limitaciones del Enfoque Morfológico

Ventajas:

- **Interpretabilidad:** Reglas de decisión explícitas y comprensibles basadas en características físicas medibles
- **Eficiencia computacional:** Cálculos geométricos y estadísticos simples, ejecución en tiempo real en hardware embebido
- **Pocos datos requeridos:** Decenas a cientos de muestras por clase, vs. miles requeridos por aprendizaje profundo
- **Control explícito:** Ajuste directo de umbrales y pesos según requerimientos operativos
- **Depuración sencilla:** Fallas identificables mediante inspección de valores de descriptores
- **Sin dependencias externas:** No requiere frameworks de deep learning ni GPUs

Limitaciones:

Proyecto Final de Estudios: Robot cosechador automático

- **Diseño manual de características:** Requiere conocimiento experto del dominio para identificar descriptores relevantes
- **Limitada a características medibles:** No captura patrones visuales complejos o texturas sutiles
- **Sensibilidad a condiciones:** Variaciones significativas de iluminación o perspectiva pueden afectar descriptores
- **Separabilidad lineal:** Dificultad con clases no linealmente separables en el espacio de características
- **Escalabilidad a muchas clases:** Desempeño puede degradar con ¿10 clases

Aplicabilidad en Agricultura de Precisión

El enfoque morfológico resulta particularmente adecuado para clasificación de cultivos cuando:

1. Existe diferenciación clara por tamaño, forma o color (plantines vs. plantas maduras vs. tubos vacíos)
2. El entorno es controlado (cultivo hidropónico con iluminación artificial)
3. Se requiere operación en hardware con recursos limitados (Raspberry Pi, microcontroladores)
4. La interpretabilidad del sistema es crítica para validación agronómica
5. El número de clases es limitado (2-5 categorías)
6. Se dispone de conocimiento experto para diseñar descriptores relevantes

La combinación con análisis estadístico robusto permite alcanzar exactitudes ¿90 % en escenarios donde las características geométricas y cromáticas son discriminativas, sin la complejidad y requisitos computacionales de métodos de aprendizaje profundo.

Técnicas Adicionales para Robustez

Normalización de Descriptores

Antes de calcular distancias, normalizar descriptores al rango $[0, 1]$ evita que características con rangos grandes dominen:

$$d_{norm} = \frac{d - d_{min}}{d_{max} - d_{min}} \quad (2.85)$$

donde d_{min} y d_{max} son los valores mínimo y máximo observados en el dataset de entrenamiento.

Umbral de Rechazo

Para evitar clasificaciones erróneas en objetos anómalos, se establece un umbral máximo de distancia:

$$\text{Clase}(\mathbf{x}) = \begin{cases} \arg \min_i D_i(\mathbf{x}) & \text{si } D_{min} < T_{rechazo} \\ \text{DESCONOCIDO} & \text{si } D_{min} \geq T_{rechazo} \end{cases} \quad (2.86)$$

Objetos con $D_{min} \geq T_{rechazo}$ son rechazados como no pertenecientes a ninguna clase conocida.

Actualización de Prototipos

En aplicaciones donde las características de las clases evolucionan (ej. crecimiento de plantas), los prototipos pueden actualizarse incrementalmente:

$$\boldsymbol{\mu}_i^{nuevo} = \alpha \cdot \boldsymbol{\mu}_i^{viejo} + (1 - \alpha) \cdot \mathbf{x}_{nuevo} \quad (2.87)$$

donde $\alpha \in [0, 0.9, 0.99]$ controla la tasa de adaptación.

2.2.4. Métricas de Evaluación para Sistemas de Clasificación

La evaluación cuantitativa del desempeño de sistemas de clasificación requiere métricas robustas que capturen diferentes aspectos de la calidad predictiva. En aplicaciones de agricultura de precisión, donde decisiones incorrectas pueden impactar la productividad del cultivo, resulta fundamental comprender las fortalezas y limitaciones de cada métrica.

Matriz de Confusión

La matriz de confusión constituye la representación fundamental del desempeño de un clasificador, tabulando predicciones versus etiquetas verdaderas. Para un problema de C clases, la matriz de confusión $\mathbf{M} \in \mathbb{R}^{C \times C}$ se define como:

$$M_{ij} = \sum_{k=1}^N \mathbb{I}[y_k = i \wedge \hat{y}_k = j] \quad (2.88)$$

donde y_k es la clase verdadera de la muestra k , \hat{y}_k es la clase predicha, y $\mathbb{I}[\cdot]$ es la función indicadora.

Para clasificación binaria, la matriz se simplifica a:

$$\mathbf{M} = \begin{bmatrix} \text{TN} & \text{FP} \\ \text{FN} & \text{TP} \end{bmatrix} \quad (2.89)$$

donde:

- **TP (True Positives):** Casos positivos correctamente clasificados
- **TN (True Negatives):** Casos negativos correctamente clasificados
- **FP (False Positives):** Casos negativos incorrectamente clasificados como positivos (Error Tipo I)
- **FN (False Negatives):** Casos positivos incorrectamente clasificados como negativos (Error Tipo II)

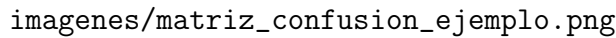
Exactitud (Accuracy)

La exactitud mide la proporción global de predicciones correctas:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{\sum_{i=1}^C M_{ii}}{N} \quad (2.90)$$

Ventajas:

- Interpretación intuitiva: porcentaje de aciertos
- Métrica estándar para comparación entre modelos
- Apropiaada cuando todas las clases tienen igual importancia



imagenes/matriz_confusion_ejemplo.png

Figura 2.1: Matriz de confusión para clasificación multiclase mostrando patrones de error

Limitaciones Críticas

La exactitud resulta engañosa en datasets desbalanceados. Considere un clasificador que siempre predice la clase mayoritaria en un dataset con 95 % de negativos:

$$\text{Accuracy}_{naive} = \frac{0,95N}{N} = 0,95 \quad (2.91)$$

Este clasificador trivial alcanza 95 % de exactitud sin aprender ningún patrón útil. En agricultura, donde ciertas condiciones pueden ser raras pero críticas, la exactitud puede ocultar fallos graves del modelo.

Precisión (Precision)

La precisión cuantifica la proporción de predicciones positivas que son correctas:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2.92)$$

Interpretación: De todas las instancias que el modelo clasificó como positivas, ¿qué fracción realmente lo era?

Para clasificación multiclase, se calcula precisión por clase:

$$\text{Precision}_i = \frac{M_{ii}}{\sum_{j=1}^C M_{ji}} \quad (2.93)$$

La precisión es crucial cuando el costo de falsos positivos es alto. En cosecha automatizada, alta precisión garantiza que el robot no intente cosechar elementos no deseados.

Exhaustividad (Recall/Sensitivity)

El recall mide la proporción de casos positivos reales que fueron correctamente identificados:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.94)$$

Interpretación: De todas las instancias que realmente son positivas, ¿qué fracción detectó el modelo?

Para clasificación multiclase:

$$\text{Recall}_i = \frac{M_{ii}}{\sum_{j=1}^C M_{ij}} \quad (2.95)$$

El recall es fundamental cuando el costo de falsos negativos es alto. En detección de elementos para cosecha, alto recall asegura que no se pierdan productos maduros.

F-Score

El F-Score armoniza precisión y recall mediante su media armónica:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (2.96)$$

Generalización F-Beta

Cuando se desea ponderar diferentemente precisión y recall:

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}} \quad (2.97)$$

- $\beta < 1$: Favorece precisión
- $\beta = 1$: Balance equitativo (F1-Score)
- $\beta > 1$: Favorece recall

Métricas Agregadas Multiclase

Para problemas con $C > 2$ clases, las métricas por clase se agregan mediante:

Macro-Average

Promedio simple de métricas por clase:

$$\text{Metric}_{macro} = \frac{1}{C} \sum_{i=1}^C \text{Metric}_i \quad (2.98)$$

Trata todas las clases por igual, apropiado cuando cada clase tiene igual importancia independientemente de su frecuencia.

Weighted-Average

Promedio ponderado por soporte de cada clase:

$$\text{Metric}_{weighted} = \frac{1}{N} \sum_{i=1}^C n_i \cdot \text{Metric}_i \quad (2.99)$$

donde $n_i = \sum_{j=1}^C M_{ij}$ es el número de instancias verdaderas de clase i .



Figura 2.2: Trade-off entre precisión y recall para diferentes valores de umbral

Micro-Average

Calcula métricas globales agregando TP, FP, FN de todas las clases:

$$\text{Precision}_{micro} = \frac{\sum_{i=1}^C \text{TP}_i}{\sum_{i=1}^C (\text{TP}_i + \text{FP}_i)} \quad (2.100)$$

Métricas de Error Absoluto

Para clasificadores basados en descriptores continuos, es relevante evaluar el error en la estimación del descriptor mismo.

Error Absoluto Medio (MAE)

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2.101)$$

Mide la magnitud promedio de los errores sin considerar su dirección.

Error Cuadrático Medio (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (2.102)$$

Penaliza errores grandes más fuertemente que MAE debido a la elevación al cuadrado.

Error Porcentual Absoluto Medio (MAPE)

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2.103)$$

Expresa el error como porcentaje del valor verdadero, útil para comparar desempeño en magnitudes diferentes.



Figura 2.3: Distribución de errores absolutos en la estimación de descriptores

Coefficiente de Determinación

Para regresión de descriptores continuos, el coeficiente R^2 mide la proporción de varianza explicada:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (2.104)$$

donde \bar{y} es la media de los valores observados.

Interpretación:

- $R^2 = 1$: Predicción perfecta
- $R^2 = 0$: Modelo no mejor que predecir la media
- $R^2 < 0$: Modelo peor que predecir la media

Métricas de Confiabilidad

Intervalo de Confianza

Para una métrica estimada $\hat{\theta}$ con error estándar SE , el intervalo de confianza al $(1 - \alpha)100\%$ es:

$$IC_{1-\alpha} = \hat{\theta} \pm z_{\alpha/2} \cdot SE \quad (2.105)$$

Para exactitud:

$$SE_{Acc} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \quad (2.106)$$

Coefficiente de Variación

Mide la variabilidad relativa de una métrica:

$$CV = \frac{\sigma}{\mu} \times 100\% \quad (2.107)$$

Valores bajos de CV indican consistencia del clasificador entre diferentes conjuntos de prueba.

Métricas de Robustez

Estabilidad ante Perturbaciones

Evalúa degradación de desempeño bajo ruido:

$$\Delta\text{Perf} = \frac{\text{Perf}_{clean} - \text{Perf}_{noisy}}{\text{Perf}_{clean}} \times 100\% \quad (2.108)$$

Sistemas robustos exhiben $\Delta\text{Perf} < 10\%$ para niveles moderados de ruido.

Repetibilidad

Desviación estándar de la métrica en ejecuciones repetidas:

$$\sigma_{rep} = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (\text{Metric}_k - \bar{\text{Metric}})^2} \quad (2.109)$$

donde K es el número de repeticiones.

Selección de Métricas según Aplicación

Para sistemas de clasificación en agricultura de precisión, la selección de métricas debe alinearse con los objetivos operativos:

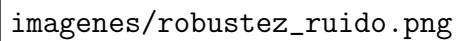
Prioridades Típicas:

1. **Alta Exhaustividad:** Minimizar productos no detectados
2. **Precisión Aceptable:** Limitar clasificaciones incorrectas
3. **Robustez:** Mantener desempeño bajo variaciones ambientales
4. **Consistencia:** Bajo coeficiente de variación entre sesiones

La métrica compuesta óptima puede definirse como:

$$\text{Score}_{total} = w_1 \cdot F_\beta + w_2 \cdot (1 - CV) + w_3 \cdot (1 - \Delta\text{Perf}) \quad (2.110)$$

donde w_i son pesos que reflejan la importancia relativa de cada aspecto, con $\sum w_i = 1$.



imagenes/robustez_ruido.png

Figura 2.4: Degradación de exactitud con incremento de ruido gaussiano

2.3. Cinemática de Robots Cartesianos

2.4. Sistemas de Transmisión Mecánica

2.5. Control de Motores Paso a Paso

3. Desarrollo del Sistema

3.1. Arquitectura General del Sistema

3.2. Modelado y Diseño Mecánico

3.2.1. Especificaciones de Diseño y Restricciones

Descripción General del Sistema

El sistema robótico diseñado para la manipulación automatizada de lechugas en cultivos hidropónicos verticales presenta una configuración cartesiana de dos ejes (XZ), con dimensiones de 3 m de ancho por 2 m de alto. El robot está equipado con un brazo robótico serie que posee un efector final tipo pinza con dos grados de libertad para la sujeción y traslado de las lechugas.

Configuración Estructural

El diseño estructural se basa en dos subsistemas de movimiento perpendiculares, cada uno optimizado según los requerimientos de carga y precisión:

El eje horizontal implementa un sistema de transmisión por correa dentada y poleas, accionado mediante motor paso a paso. La característica distintiva de este diseño es que la carga móvil se soporta mediante un perfil tipo poli-V equipado con ruedas de desplazamiento. Esta configuración permite que la correa de transmisión trabaje prácticamente sin carga de peso, funcionando únicamente como elemento de tracción, lo cual mejora significativamente la precisión del posicionamiento y reduce el desgaste del sistema.

El conjunto incluye un tensor de correa ajustable que mantiene la tensión óptima durante la operación, compensando elongaciones del material y garantizando transmisión sin deslizamiento. Además cuenta con perfiles guías, uno en cada lateral del soporte superior, ambos paralelos al perfil de soporte superior. Esto da rigidez lateral al sistema y anulando la oscilación alrededor del perfil horizontal.

El movimiento vertical se realiza mediante transmisión por tornillo de potencia, proporcionando capacidad de carga superior y autobloqueo mecánico. Para garantizar la rigidez del sistema y evitar rotaciones indeseadas del carro móvil, se incorporan dos varillas lisas de acero dispuestas en paralelo a la varilla roscada. Esta configuración de tres varillas paralelas asegura un movimiento vertical estable y preciso: la varilla roscada transmite el movimiento mientras que las dos varillas lisas actúan como guías lineales, proporcionando rigidez lateral y evitando giros del soporte, especialmente bajo cargas asimétricas.

Por otro lado, el robot cuenta con tres soportes estructurales principales fabricados mediante impresión 3D:

- **Soporte inferior:** Fijación de las varillas verticales y montaje del motor del eje Z
- **Soporte superior:** Permite el acoplamiento cinemático de los dos movimientos perpendiculares (horizontal y vertical), actuando como elemento de unión entre ambos subsistemas
- **Soporte medio:** Sostiene el brazo robótico y la cámara y se mueve cuando gira la varilla.

Proyecto Final de Estudios: Robot cosechador automático

Cargas del Sistema

Las cargas consideradas en el diseño se distribuyen según su función operativa:

Cargas del eje horizontal: Es la suma del peso de las varillas, la varilla roscada, la carga, los soportes y el brazo. Considerando que aproximadamente toda la carga la sostiene el perfil superior, se tiene que la carga aproximada que debe mover el motor es de aproximadamente 2kg

Cargas del eje vertical: Es la suma del peso de la lechuga y el brazo. Además el motor debe vencer la fricción que se genera entre la tuerca y la varilla.

Especificaciones Cinemáticas

Las especificaciones de velocidad, aceleración y resolución de los actuadores se detallan en la Tabla 3.1.

Cuadro 3.1: Especificaciones cinemáticas de los ejes de movimiento

Parámetro	Eje Horizontal	Eje Vertical
Velocidad máxima $\frac{mm}{s}$	250	75
Velocidad mínima $\frac{mm}{s}$	12.5	2.5
Aceleración $\frac{mm}{s^2}$	187.5	45
Avance por revolución $\frac{mm}{rev}$	40	8
Resolución $\frac{pasos}{mm}$	40	200
Micropasos	8	8

La diferencia en resolución entre ambos ejes responde a los requisitos operativos: el eje horizontal requiere mayor velocidad de desplazamiento con precisión moderada, mientras que el eje vertical necesita mayor precisión de posicionamiento con velocidades menores debido a las operaciones de elevación de carga.

3.2.2. Análisis Cinemático del Sistema

3.2.3. Diseño Estructural

3.2.4. Sistema de Movimiento Horizontal

3.2.5. Sistema de Movimiento Vertical

3.2.6. Brazo Robótico

3.2.7. Modelado CAD y Fabricación

3.3. Sistema de Control de Bajo Nivel (Nivel Regulatorio)

3.3.1. Arquitectura del Nivel Regulatorio

3.3.2. Hardware de Control - Arduino Mega 2560

3.3.3. Selección y Dimensionamiento de Actuadores

3.3.4. Sensores de Seguridad

3.3.5. Control de Movimiento

3.3.6. Protocolo de Comunicación UART

3.4. Sistema de Supervisión y Alta Gestión (Nivel Supervisor)

3.4.1. Arquitectura del Nivel Supervisor

3.4.2. Hardware Supervisor - Raspberry Pi 4

3.4.3. Máquina de Estados Supervisora

3.4.4. Coordinación con Nivel Regulatorio

3.5. Inteligencia Artificial y Visión por Computadora

3.5.1. Arquitectura y Pipeline de Procesamiento

Arquitectura Modular del Sistema

El sistema de inteligencia artificial implementado se estructura en cuatro módulos funcionales especializados que operan de manera coordinada para lograr la autonomía del robot cosechador. Esta arquitectura modular permite la separación de responsabilidades, facilita el mantenimiento y posibilita la evolución independiente de cada componente.


El primer módulo, denominado **Sistema de Posicionamiento Visual**, tiene como función detectar desviaciones respecto a la posición ideal mediante el análisis de marcadores visuales de referencia. Este módulo calcula correcciones en tiempo real que compensan las holguras mecánicas acumulativas del sistema de movimiento, garantizando precisión submilimétrica en el alineamiento frente a cada estación de cultivo.

Proyecto Final de Estudios: Robot cosechador automático

El segundo módulo corresponde al **Sistema de Clasificación de Cultivos**, responsable de determinar la presencia o ausencia de lechuga en cada estación mediante análisis morfológico de características extraídas de imágenes. La clasificación se basa en umbrales estadísticos derivados de un análisis exhaustivo de muestras representativas del entorno operativo.

El tercer módulo, **Sistema de Mapeo Autónomo**, construye una representación espacial completa del entorno de cultivo mediante exploración sistemática. Este módulo registra la posición y estado de cada estación, generando una base de datos espacial que sirve como entrada para la planificación de operaciones de cosecha.

Finalmente, el cuarto módulo de **Optimización de Trayectorias** procesa la información del mapa construido para generar rutas que minimizan el tiempo total de operación. Este módulo adapta el problema clásico del viajante a las restricciones geométricas y cinemáticas específicas del robot cartesiano.



imagenes/arquitectura_modular_ia.png

Figura 3.1: Arquitectura modular del sistema de inteligencia artificial mostrando la interconexión entre los cuatro módulos principales y el flujo de información

Flujo de Información entre Módulos

La información fluye siguiendo una jerarquía de control bien definida. El módulo de mapeo identifica las estaciones objetivo que requieren atención. Para cada estación identificada, el módulo de

posicionamiento ejecuta las correcciones visuales necesarias para garantizar el alineamiento preciso. Una vez posicionado correctamente, el módulo de clasificación determina el estado del cultivo. Finalmente, basándose en esta información, el sistema supervisor toma la decisión de ejecutar la cosecha o avanzar a la siguiente estación. Este flujo puede expresarse mediante la secuencia:

$$\text{Mapeo} \rightarrow \text{Posicionamiento} \rightarrow \text{Clasificación} \rightarrow \text{Decisión} \quad (3.1)$$

Cada módulo expone interfaces funcionales bien definidas que permiten la invocación de operaciones específicas y el intercambio de datos estructurados. El módulo de posicionamiento recibe como entrada la imagen capturada y el eje a corregir, retornando la desviación calculada en milímetros. El módulo de clasificación procesa una imagen y retorna un valor booleano indicando presencia o ausencia de lechuga junto con un índice de confianza. El módulo de mapeo ejecuta la exploración completa y retorna una matriz que contiene coordenadas y estados de todas las estaciones. El módulo de trayectorias recibe esta matriz como entrada y genera una lista ordenada de coordenadas que define la secuencia óptima de visita.

Selección de Plataforma Computacional

La implementación del sistema de inteligencia artificial se realizó sobre una Raspberry Pi 4 Model B equipada con procesador ARM Cortex-A72 de cuatro núcleos operando a 1.5 GHz y 4 GB de memoria RAM tipo LPDDR4. Esta plataforma fue seleccionada por su capacidad de ejecutar algoritmos de visión por computadora en tiempo real manteniendo un consumo energético reducido, compatible con las restricciones del sistema embarcado del robot.

Las bibliotecas de software empleadas incluyen OpenCV versión 4.5 para procesamiento de imágenes y detección de contornos, NumPy versión 1.21 para operaciones matriciales de alto rendimiento necesarias en la calibración espacial, y PySerial para la comunicación mediante protocolo UART con el nivel regulatorio. La elección de Python como lenguaje de implementación se fundamenta en su amplio ecosistema de bibliotecas científicas, la velocidad de desarrollo y prototipado, y un rendimiento suficiente para las frecuencias de operación requeridas por el sistema.

El sistema operativo Raspberry Pi OS, basado en Debian, proporciona soporte nativo para los controladores de cámara mediante la interfaz Video4Linux2, permitiendo la captura de imágenes sin configuración adicional. La arquitectura ARM del procesador incluye extensiones SIMD que aceleran las operaciones vectoriales intensivas empleadas en el procesamiento de imágenes.

Pipeline General de Procesamiento

El procesamiento de imágenes sigue un pipeline estructurado en siete etapas secuenciales que transforman las imágenes crudas capturadas en información accionable para el control del robot. Cada etapa aplica transformaciones específicas que progresivamente refinan la información hasta extraer las características relevantes para la toma de decisiones.

La primera etapa corresponde a la **adquisición de imágenes** RGB con resolución de 1920×1080 píxeles mediante la cámara USB. Esta etapa presenta una latencia aproximada de 50 milisegundos que incluye el tiempo de exposición del sensor, la lectura de los datos del array de píxeles y la transferencia mediante interfaz USB 2.0 hacia la memoria del sistema.

La segunda etapa realiza el **preprocesamiento** mediante la conversión del espacio de color RGB al espacio HSV. Esta transformación, fundamentada en la teoría presentada en el marco teórico (Sección 2.2.1), separa la información cromática (matiz y saturación) de la información de luminosidad (valor), proporcionando robustez ante variaciones de iluminación ambiental. La conversión se ejecuta aplicando las ecuaciones de transformación estándar RGB-HSV sobre cada píxel de la imagen. El tiempo de procesamiento de esta operación es de aproximadamente 15 milisegundos, aprovechando las optimizaciones vectoriales de la biblioteca OpenCV.

Proyecto Final de Estudios: Robot cosechador automático

La tercera etapa corresponde a la **segmentación**, donde se aplican operaciones de umbralización específicas según el módulo activo. Para el módulo de posicionamiento se extrae el canal V y se aplica umbralización inversa con valor de referencia de 50, destacando las regiones oscuras correspondientes a las cintas de referencia. Para el módulo de clasificación se emplea segmentación por rango cromático en los tres canales HSV, definiendo límites inferior y superior que aíslan las tonalidades verdes correspondientes a vegetación. Esta etapa requiere 20 milisegundos de procesamiento.

La cuarta etapa implementa **refinamiento morfológico** mediante operaciones de cierre y apertura, aplicando los fundamentos teóricos descritos en la Sección 2.2.1 del marco teórico. La operación de cierre elimina pequeños huecos dentro de las regiones de interés, mientras que la apertura elimina píxeles aislados correspondientes a ruido. Ambas operaciones emplean un elemento estructurante rectangular de 3×3 píxeles. El costo computacional total de estas operaciones morfológicas es de 25 milisegundos.

La quinta etapa ejecuta la **detección de contornos** mediante el algoritmo de Suzuki-Abe, identificando las fronteras de las regiones segmentadas. Este algoritmo, cuya fundamentación matemática se presentó en la Sección 2.2.2, construye una representación jerárquica de los contornos detectados. Se emplea el modo de extracción de contornos externos únicamente, ya que las estructuras de interés no presentan huecos internos relevantes. La aproximación de contornos se realiza mediante compresión de puntos colineales, reduciendo el uso de memoria. Esta operación requiere aproximadamente 30 milisegundos de procesamiento.

La sexta etapa calcula **descriptores geométricos** de los contornos detectados. Se calculan el área mediante conteo de píxeles interiores, el centroide mediante momentos de imagen de orden cero y uno, el perímetro mediante suma de distancias euclidianas entre puntos consecutivos, y el rectángulo delimitador que encierra completamente al contorno. Estos descriptores constituyen las características cuantitativas empleadas para la toma de decisiones en los módulos de posicionamiento y clasificación. El tiempo de cálculo es de aproximadamente 10 milisegundos.

Finalmente, la séptima etapa genera el **comando de acción** correspondiente basándose en las características extraídas. Para el módulo de posicionamiento, esto corresponde al cálculo de la desviación en píxeles respecto al centro de la imagen y su posterior conversión a milímetros mediante los coeficientes de calibración. Para el módulo de clasificación, corresponde a la evaluación del área del contorno principal contra el umbral estadístico establecido, resultando en una decisión binaria de presencia o ausencia de lechuga. Esta etapa final requiere menos de 5 milisegundos.

El tiempo total de procesamiento del pipeline completo se obtiene mediante la suma de las latencias individuales:

$$T_{total} = T_{adq} + T_{prep} + T_{seg} + T_{mor} + T_{cont} + T_{desc} + T_{dec} \quad (3.2)$$

$$T_{total} = 50 + 15 + 20 + 25 + 30 + 10 + 5 = 155 \text{ ms} \quad (3.3)$$

Esta latencia total permite una frecuencia de operación de:

$$f = \frac{1}{T_{total}} = \frac{1}{0,155} \approx 6,5 \text{ Hz} \quad (3.4)$$

Esta frecuencia es suficiente para el control en tiempo real considerando que las velocidades de movimiento del robot se encuentran en el rango de 5 a 10 cm/s, lo que implica desplazamientos de 7.75 a 15.5 mm entre capturas consecutivas.

El pipeline implementa un diseño adaptable que modifica su configuración según el módulo activo. Para el módulo de posicionamiento se utiliza únicamente el canal V del espacio HSV con umbralización inversa, enfocándose en la detección de elementos de alto contraste. Para el módulo de clasificación se emplean los tres canales HSV con segmentación por rango cromático, permitiendo la discriminación de vegetación verde. Esta adaptabilidad permite optimizar el procesamiento según los

requerimientos específicos de cada tarea, manteniendo la latencia dentro de límites aceptables para el control en tiempo real.

Sistema de Adquisición Visual

El hardware de adquisición de imágenes constituye el elemento sensorial primario del sistema de inteligencia artificial. La selección del dispositivo de captura se basó en criterios de resolución, frecuencia de muestreo, compatibilidad con la plataforma de procesamiento y costo.

Se implementó una cámara USB Logitech C920 HD Pro que proporciona captura de imágenes con resolución máxima de 1920×1080 píxeles a una tasa de 30 cuadros por segundo. El sensor de imagen tipo CMOS de 1/2.7 pulgadas presenta sensibilidad adecuada para las condiciones de iluminación del invernadero, con valores típicos de 800 a 1200 lux proporcionados por iluminación LED blanca de 5000K instalada en el techo de la estructura.

El campo de visión de 78 grados en diagonal permite capturar un área aproximada de 280×160 milímetros a la distancia de trabajo nominal de 200 milímetros. Esta configuración óptica resulta en una resolución espacial de 0.146 mm/píxel en dirección horizontal y 0.148 mm/píxel en dirección vertical, calculadas como el cociente entre las dimensiones físicas del área capturada y la resolución en píxeles correspondiente.

La cámara incorpora sistema de enfoque automático que ajusta la posición de las lentes para mantener la nitidez de la imagen cuando varía la distancia de trabajo. Este mecanismo es crítico dado que el robot opera en un rango de distancias de 150 a 250 milímetros durante las diferentes fases de operación. El tiempo de convergencia del autofocus es inferior a 2 segundos cuando se produce un cambio significativo de distancia.

El montaje de la cámara se realizó mediante un soporte diseñado específicamente e impreso en material PLA mediante manufactura aditiva. El soporte se fija al extremo del brazo robótico y permite ajuste angular de ± 15 grados en los ejes de cabeceo y guiñada. Esta capacidad de ajuste es necesaria para compensar tolerancias de ensamblaje y optimizar la orientación del campo de visión respecto al plano de cultivo.

La interfaz entre la cámara y la Raspberry Pi se establece mediante conexión USB 2.0, que proporciona ancho de banda suficiente para la transferencia de imágenes sin compresión a la tasa de captura requerida. El controlador Video4Linux2, incluido nativamente en el sistema operativo, gestiona la comunicación con el dispositivo y expone las funciones de captura mediante una interfaz de programación estándar.

La latencia total del sistema de adquisición, medida desde la solicitud de captura hasta la disponibilidad de la imagen en memoria, comprende cuatro componentes: tiempo de exposición del sensor (típicamente 8 a 15 milisegundos dependiendo de las condiciones de iluminación), tiempo de lectura del array de píxeles (aproximadamente 8 milisegundos), tiempo de transferencia USB (15 milisegundos para una imagen de 6.2 megabytes), y tiempo de decodificación en el procesador host (10 milisegundos). La latencia total resultante es de aproximadamente 45 a 50 milisegundos.

Preprocesamiento de Imágenes

Las imágenes capturadas requieren transformaciones preliminares que optimizan su procesamiento posterior y mejoran la robustez del sistema ante variaciones ambientales. El preprocesamiento constituye la etapa de acondicionamiento de señal del pipeline de visión artificial.

La primera operación de preprocesamiento consiste en la conversión del espacio de color RGB nativo de la cámara al espacio HSV. Esta transformación aplica las ecuaciones de conversión estándar presentadas en el marco teórico (Sección 2.2.1), calculando para cada píxel los valores de matiz, saturación y valor a partir de las componentes rojo, verde y azul. La separación de información

cromática e información de luminosidad que proporciona esta representación resulta fundamental para la robustez del sistema.

El módulo de posicionamiento visual explota esta separación extrayendo únicamente el canal V, que representa el brillo o luminosidad de cada píxel independientemente de su color. Esta estrategia proporciona invariancia ante cambios en la tonalidad de la iluminación, permitiendo la detección de cintas negras de referencia incluso cuando el espectro de la fuente luminosa varía. El canal V se extrae mediante indexación directa del tercer plano de la imagen transformada.

El módulo de clasificación de cultivos, por el contrario, emplea los tres canales simultáneamente para realizar segmentación por rango cromático. Esta operación evalúa para cada píxel si sus valores de matiz, saturación y valor se encuentran dentro de límites predefinidos que corresponden a las tonalidades verdes características de la vegetación. Los píxeles que cumplen esta condición se marcan en una máscara binaria, aislando efectivamente las regiones de interés del resto de la escena.

Para optimizar el rendimiento computacional se implementa recorte adaptativo de región de interés. Este mecanismo limita el procesamiento a una ventana rectangular centrada en la posición esperada del elemento a detectar, reduciendo significativamente el número de píxeles que deben procesarse. Las dimensiones de la región de interés se determinan en función de la posición actual del robot y del conocimiento a priori de la geometría del sistema. Las coordenadas de los límites de la región se calculan mediante:

$$x_{min} = \max(0, x_{centro} - w_{ROI}/2) \quad (3.5)$$

$$x_{max} = \min(W_{imagen}, x_{centro} + w_{ROI}/2) \quad (3.6)$$

$$y_{min} = \max(0, y_{centro} - h_{ROI}/2) \quad (3.7)$$

$$y_{max} = \min(H_{imagen}, y_{centro} + h_{ROI}/2) \quad (3.8)$$

donde W_{imagen} y H_{imagen} son las dimensiones totales de la imagen capturada, (x_{centro}, y_{centro}) es la posición estimada del elemento de interés, y w_{ROI} y h_{ROI} son las dimensiones de la región de interés. Las funciones de máximo y mínimo garantizan que los límites no excedan los bordes de la imagen.

Esta estrategia reduce el área de procesamiento en aproximadamente 65 por ciento, mejorando proporcionalmente los tiempos de respuesta del sistema. El recorte se ejecuta mediante operaciones de indexación de arrays que no implican copia de datos, resultando en un costo computacional despreciable.

Segmentación por Umbralización

La segmentación constituye la operación fundamental que separa las regiones de interés del fondo de la imagen. Se implementan dos estrategias de umbralización adaptadas a los requerimientos específicos de cada módulo.

Para el módulo de posicionamiento visual se aplica umbralización inversa sobre el canal V del espacio HSV. Esta operación compara el valor de brillo de cada píxel contra un umbral de referencia, asignando valor máximo (255) a los píxeles cuyo brillo es inferior al umbral, y valor nulo (0) a los píxeles con brillo superior. Matemáticamente:

$$I_{bin}(x, y) = \begin{cases} 255 & \text{si } I_V(x, y) < T \\ 0 & \text{si } I_V(x, y) \geq T \end{cases} \quad (3.9)$$

El valor de umbral $T = 50$ se determinó empíricamente mediante análisis del histograma de intensidades en imágenes representativas del entorno operativo. Este valor maximiza la separación entre las cintas negras de referencia y el fondo blanco del sistema hidropónico, minimizando simultáneamente la tasa de falsos positivos.

Para el módulo de clasificación se emplea segmentación por rango en el espacio tridimensional HSV. Cada píxel se evalúa contra límites inferior y superior en los tres canales, asignándose a la región de interés únicamente si todas las componentes se encuentran dentro del rango especificado:

$$M(x, y) = \begin{cases} 255 & \text{si } H_{min} \leq H(x, y) \leq H_{max} \\ & \wedge S_{min} \leq S(x, y) \leq S_{max} \\ & \wedge V_{min} \leq V(x, y) \leq V_{max} \\ 0 & \text{en caso contrario} \end{cases} \quad (3.10)$$

Los límites se establecieron como $H_{min} = 25^\circ$, $H_{max} = 85^\circ$ para cubrir el espectro de verdes desde tonalidades amarillentas hasta azuladas; $S_{min} = 40$ para eliminar colores desaturados del fondo que podrían presentar componente verde débil; y $V_{min} = 40$ para descartar sombras muy oscuras. Los límites superiores de saturación y valor se fijan en sus valores máximos (255).

Refinamiento Morfológico

Las máscaras binarias resultantes de la segmentación contienen típicamente ruido aislado y discontinuidades en las regiones de interés debido a variaciones locales de iluminación, reflexiones especulares y limitaciones del sensor. Se aplica una secuencia de operaciones morfológicas para refinar estas máscaras.

La primera operación corresponde al cierre morfológico, que elimina pequeños huecos dentro de las regiones y conecta componentes próximos. Matemáticamente, el cierre se define como una dilatación seguida de una erosión:

$$M_{cierre} = (M \oplus K) \ominus K \quad (3.11)$$

donde M es la máscara binaria de entrada, K es el elemento estructurante, \oplus denota dilatación y \ominus denota erosión. La dilatación expande las regiones blancas, rellenando huecos pequeños, mientras que la erosión subsecuente restaura aproximadamente el tamaño original de las regiones.

La segunda operación corresponde a la apertura morfológica, que elimina píxeles aislados y pequeñas protuberancias. Se define como una erosión seguida de una dilatación:

$$M_{apertura} = (M_{cierre} \ominus K) \oplus K \quad (3.12)$$

La erosión elimina estructuras pequeñas que no pueden contener completamente el elemento estructurante, y la dilatación subsecuente restaura el tamaño de las estructuras remanentes.

El elemento estructurante empleado es una matriz rectangular de 3×3 píxeles con todos sus elementos igual a uno. Este tamaño se seleccionó como compromiso entre capacidad de eliminación de ruido y preservación de detalles relevantes. Elementos estructurantes mayores eliminarían ruido más efectivamente pero podrían degradar características geométricas importantes de las regiones.

El resultado de estas operaciones morfológicas es una máscara binaria refinada donde las regiones de interés presentan fronteras suaves y conectividad mejorada, facilitando la subsecuente detección de contornos.

Detección y Análisis de Contornos

Sobre la máscara refinada se ejecuta el algoritmo de detección de contornos de Suzuki-Abe, cuyo fundamento teórico se presentó en la Sección 2.2.2 del marco teórico. Este algoritmo recorre la imagen binaria identificando las fronteras entre regiones blancas y negras, construyendo una representación vectorial de cada contorno como secuencia ordenada de coordenadas de píxeles.

Se emplea el modo de extracción de contornos externos únicamente, que detecta solamente el contorno exterior de cada región conexas ignorando posibles huecos internos. Esta configuración es

apropiada dado que los objetos de interés (lechugas y cintas de referencia) no presentan estructuras anidadas relevantes para la aplicación. La aproximación de contornos se realiza mediante compresión de segmentos colineales, donde secuencias de puntos que forman líneas rectas se representan únicamente por sus extremos. Esta estrategia reduce significativamente el uso de memoria y acelera operaciones posteriores sobre los contornos.

Los contornos detectados se someten a filtrado para eliminar aquellos que no corresponden a objetos de interés. El criterio principal de filtrado es el área, calculada como el número de píxeles encerrados por el contorno. Para el módulo de posicionamiento se descartan contornos con área inferior a 500 píxeles, mientras que para el módulo de clasificación el umbral es de 5000 píxeles. Estos valores se establecieron considerando el tamaño esperado de los objetos de interés a la distancia de trabajo nominal.

Para cada contorno que supera el filtrado se calculan descriptores geométricos que cuantifican sus características. El área se obtiene mediante conteo de píxeles interiores empleando el algoritmo de relleno por escaneo de líneas. El centroide se calcula mediante los momentos de imagen de orden cero y uno:

$$c_x = \frac{M_{10}}{M_{00}}, \quad c_y = \frac{M_{01}}{M_{00}} \quad (3.13)$$

donde M_{ij} representa el momento de orden $i + j$ definido como:

$$M_{ij} = \sum_{(x,y) \in \text{Región}} x^i y^j \quad (3.14)$$

El perímetro se calcula como la suma de distancias euclidianas entre puntos consecutivos del contorno. El rectángulo delimitador se determina identificando las coordenadas mínimas y máximas en ambas direcciones.

Cuando múltiples contornos cumplen los criterios de filtrado, se implementa un sistema de selección del contorno principal. Para el módulo de posicionamiento se emplea scoring ponderado que considera simultáneamente el área del contorno, su posición relativa al centro de la imagen, y la calidad de su región basal. Para el módulo de clasificación se selecciona simplemente el contorno de mayor área, bajo la hipótesis de que la lechuga constituye el objeto verde dominante en la escena.

El tiempo total de procesamiento de la etapa de detección y análisis de contornos es de aproximadamente 40 milisegundos, incluyendo la detección, el filtrado y el cálculo de descriptores. Este tiempo es compatible con los requerimientos de latencia del sistema de control en tiempo real.

3.5.2. Sistema de Posicionamiento Visual

Fundamento del Sistema de Posicionamiento Visual

El sistema mecánico del robot presenta holguras acumulativas que generan errores de posicionamiento de hasta ± 5 milímetros respecto a la posición comandada. Estos errores resultan incompatibles con los requerimientos de precisión para las operaciones de cosecha, que exigen tolerancias inferiores a ± 2 milímetros. Para compensar estas desviaciones se implementó un sistema de corrección visual basado en la detección de marcadores de referencia.

Los marcadores consisten en cintas adhesivas negras de 18 milímetros de ancho adheridas sobre la superficie de los tubos de PVC blanco del sistema hidropónico. Esta configuración proporciona un contraste óptico superior a 0.85 entre el elemento oscuro y el fondo claro, facilitando su detección mediante técnicas de procesamiento de imágenes. La selección de cintas adhesivas como elemento de marcado se fundamentó en su bajo costo de implementación (inferior a 5 dólares estadounidenses por metro lineal), su robustez ante las condiciones de humedad y temperatura del invernadero, y la facilidad de instalación sin modificaciones estructurales del sistema.

Las cintas se disponen en dos orientaciones: cintas horizontales que proporcionan referencia para corrección en el eje vertical, y cintas verticales que proporcionan referencia para corrección en el eje horizontal. Esta configuración dual permite la corrección independiente en ambos ejes de movimiento del robot.

Metodología de Detección

El algoritmo de detección de cintas explota el alto contraste entre el marcador negro y el fondo claro mediante procesamiento en el canal de valor (V) del espacio HSV. Esta elección se fundamenta en la invariancia del canal V ante cambios en la tonalidad cromática de la iluminación, propiedad demostrada en el marco teórico (Sección 2.2.1).

El proceso inicia con la captura de una imagen RGB de la región donde se espera encontrar el marcador. La imagen se transforma al espacio HSV aplicando las ecuaciones de conversión estándar, y se extrae el canal V que representa el brillo de cada píxel. Sobre este canal se aplica umbralización inversa con valor de referencia de 50, operación que asigna valor máximo (blanco) a los píxeles oscuros y valor nulo (negro) a los píxeles claros. Matemáticamente:

$$I_{bin}(x, y) = \begin{cases} 255 & \text{si } I_V(x, y) < 50 \\ 0 & \text{si } I_V(x, y) \geq 50 \end{cases} \quad (3.15)$$

El valor de umbral 50 se determinó empíricamente mediante análisis del histograma de intensidades en múltiples imágenes representativas, identificando el punto que maximiza la separación entre la distribución de intensidades de las cintas negras y la distribución de intensidades del fondo blanco.

Sobre la imagen binaria resultante se ejecuta el algoritmo de detección de contornos de Suzuki-Abe, obteniendo las fronteras de todas las regiones oscuras detectadas. Los contornos se filtran aplicando un criterio de área mínima de 500 píxeles, eliminando elementos espurios correspondientes a ruido o pequeñas sombras. Este umbral se estableció considerando que una cinta de 18 milímetros de ancho a 200 milímetros de distancia subtiende aproximadamente 120 píxeles de ancho en la imagen, resultando en áreas típicas superiores a 3000 píxeles para segmentos de cinta de longitud razonable.

Evaluación de Calidad del Contorno

Los contornos que superan el filtrado de área se someten a evaluación de calidad para discriminar entre cintas de referencia genuinas y otros elementos oscuros presentes en la escena (como plantas o vasos). La evaluación se basa en el análisis de la región basal del contorno, definida como el 10 por ciento inferior de su rectángulo delimitador.

Para cada contorno candidato se calcula su rectángulo delimitador (x, y, w, h) , donde (x, y) representa la esquina superior izquierda, w el ancho y h la altura. La región basal se define entonces como:

$$R_{base} = \{(x', y') : x \leq x' < x + w, y + 0,9h \leq y' < y + h\} \quad (3.16)$$

Se calcula la fracción de píxeles blancos en esta región respecto al total de píxeles de la región:

$$Q_{base} = \frac{1}{|R_{base}|} \sum_{(x,y) \in R_{base}} \frac{I_{bin}(x, y)}{255} \quad (3.17)$$

donde $|R_{base}|$ denota el número de píxeles en la región. Un valor Q_{base} cercano a 1 indica que la región basal del contorno es consistentemente oscura, característica de las cintas de referencia cuyo borde inferior es nítido y bien definido. Valores inferiores indican contornos de objetos con bases irregulares o discontinuas, como plantas con hojas que no presentan un borde horizontal claro.

Esta métrica resulta efectiva para distinguir cintas de otros elementos oscuros: las cintas genuinas presentan típicamente $Q_{base} > 0,8$, mientras que plantas u otros objetos tienen $Q_{base} < 0,5$. Se establece un umbral de aceptación de 0.7 como compromiso entre sensibilidad y especificidad.

Sistema de Scoring Multicriteria

Cuando múltiples contornos cumplen los criterios de área mínima y calidad basal, se implementa un sistema de scoring ponderado que considera tres factores para seleccionar el mejor candidato.

El primer factor evalúa el área del contorno normalizada respecto a los límites mínimo y máximo esperados:

$$S_{\text{área}} = \frac{A - A_{\min}}{A_{\max} - A_{\min}} \quad (3.18)$$

donde A es el área del contorno, $A_{\min} = 500$ píxeles y $A_{\max} = 50000$ píxeles. Este factor favorece contornos con áreas intermedias, descartando elementos excesivamente pequeños (ruido) o excesivamente grandes (probablemente no corresponden a una cinta individual).

El segundo factor evalúa la posición del centroide del contorno respecto al centro de la imagen:

$$S_{\text{posición}} = 1 - \frac{|c_x - x_{\text{centro}}|}{W_{\text{imagen}}/2} \quad (3.19)$$

donde c_x es la coordenada horizontal del centroide, x_{centro} es la coordenada horizontal del centro de la imagen, y W_{imagen} es el ancho total de la imagen. Este factor favorece contornos cercanos al centro, bajo la hipótesis de que el sistema de control posiciona aproximadamente el robot frente a la cinta objetivo, resultando en desviaciones pequeñas.

El tercer factor corresponde directamente a la calidad basal Q_{base} calculada previamente, que favorece contornos con bordes inferiores bien definidos.

El score total se calcula como combinación lineal ponderada de estos tres factores:

$$S_{\text{total}} = w_1 \cdot S_{\text{área}} + w_2 \cdot S_{\text{posición}} + w_3 \cdot Q_{base} \quad (3.20)$$

Los pesos se establecieron como $w_1 = 0,3$, $w_2 = 0,4$ y $w_3 = 0,3$, otorgando mayor importancia a la posición dado que las desviaciones típicas son inferiores a 50 milímetros, resultando en centroides próximos al centro de la imagen. El contorno con mayor score total se selecciona como la cinta de referencia.

Cálculo de Desviación

Una vez identificado el contorno óptimo se procede al cálculo de la desviación espacial. Se determina el centroide del contorno mediante sus momentos de imagen:

$$c_x = \frac{M_{10}}{M_{00}}, \quad c_y = \frac{M_{01}}{M_{00}} \quad (3.21)$$

La desviación en píxeles se calcula como la diferencia entre las coordenadas del centroide y las coordenadas del centro de la imagen:

$$\Delta x_{px} = c_x - \frac{W_{\text{imagen}}}{2} \quad (3.22)$$

$$\Delta y_{py} = c_y - \frac{H_{\text{imagen}}}{2} \quad (3.23)$$

Para una imagen de 1920×1080 píxeles, el centro se ubica en (960, 540). Un centroide detectado en (1100, 540) resultaría en $\Delta x_{px} = 140$ píxeles, indicando que el robot se encuentra desplazado 140 píxeles hacia la izquierda respecto a la posición ideal.

Estas desviaciones en píxeles se convierten posteriormente a milímetros mediante los coeficientes de calibración, proceso que se detalla en la siguiente sección.

Discriminación de Orientación

Para determinar si un contorno detectado corresponde a una cinta horizontal o vertical se analiza la relación de aspecto de su rectángulo delimitador. Esta relación se define como:

$$AR = \frac{w}{h} \quad (3.24)$$

donde w y h son el ancho y altura del rectángulo delimitador. Cintas horizontales, al tener su dimensión principal en dirección horizontal, presentan $AR > 2$. Cintas verticales, con su dimensión principal en dirección vertical, presentan $AR < 0,5$. Contornos con relaciones de aspecto intermedias ($0,5 \leq AR \leq 2$) se consideran ambiguos y se descartan.

Este análisis permite al sistema identificar automáticamente el tipo de cinta detectada, información necesaria para interpretar correctamente la desviación calculada. Una desviación horizontal del centroide respecto al centro indica desplazamiento en X cuando se detecta una cinta vertical, o desplazamiento en Y cuando se detecta una cinta horizontal.

Desempeño del Algoritmo

El tiempo de ejecución promedio del algoritmo completo de detección de cintas es de 95 milisegundos, desglosado en: conversión al espacio HSV (15 ms), extracción del canal V (2 ms), umbralización (8 ms), detección de contornos (30 ms), filtrado inicial (5 ms), evaluación de calidad basal (20 ms), cálculo de scores (10 ms), y determinación del centroide (5 ms).

Esta latencia permite una frecuencia de actualización de aproximadamente 10 Hz, suficiente para compensar errores de posicionamiento durante el movimiento del robot a velocidades de 5 a 10 cm/s. En pruebas de validación con 200 imágenes de cintas bajo diferentes condiciones de iluminación, el algoritmo alcanzó una tasa de detección correcta del 97.5 por ciento, con fallos principalmente en casos de iluminación extremadamente baja (inferior a 400 lux) donde el contraste se reduce significativamente.

Fundamentación de la Calibración Espacial

Las desviaciones detectadas mediante el sistema de visión se expresan inicialmente en unidades de píxeles, magnitud que carece de significado físico directo para el sistema de control de movimiento. Se requiere establecer una correspondencia entre coordenadas en el plano de la imagen y coordenadas en el espacio físico del robot. Esta transformación se logra mediante un proceso de calibración que determina los factores de conversión píxel-milímetro.

La calibración se fundamenta en la geometría proyectiva de formación de imágenes. Bajo la aproximación de modelo de cámara estenopeica, existe una relación lineal entre distancias en el plano objeto (espacio físico) y distancias en el plano imagen (espacio de píxeles), siempre que la distancia focal y la distancia objeto-cámara permanezcan constantes. Esta relación se expresa mediante:

$$d_{mm} = K \cdot d_{px} \quad (3.25)$$

donde d_{mm} es la distancia física en milímetros, d_{px} es la distancia correspondiente en píxeles, y K es el factor de conversión con unidades mm/píxel.

Metodología de Calibración

El proceso de calibración empleó un patrón de referencia con dimensiones conocidas colocado en el plano de trabajo a la distancia nominal de 200 milímetros desde el sensor de la cámara. El patrón consistió en un conjunto de marcas separadas por distancias calibradas mediante instrumentos de metrología de precisión.

Se capturaron imágenes del patrón y se identificaron las posiciones en píxeles de las marcas de referencia mediante detección de contornos. Para cada par de marcas consecutivas se registró la distancia física conocida $d_{mm,i}$ y la distancia medida en píxeles $d_{px,i}$. Se recolectaron $N = 20$ mediciones de distancias que cubrían el rango completo del espacio de trabajo.

El factor de conversión óptimo se determinó mediante el método de mínimos cuadrados, que minimiza la suma de errores cuadráticos entre las distancias físicas reales y las distancias estimadas mediante el modelo lineal. Para el caso univariable (factor sin offset), la solución se obtiene como:

$$K = \frac{\sum_{i=1}^N d_{px,i} \cdot d_{mm,i}}{\sum_{i=1}^N d_{px,i}^2} \quad (3.26)$$

Esta expresión proporciona el estimador de máxima verosimilitud del factor de conversión bajo el supuesto de errores gaussianos independientes.

Para considerar posibles offsets sistemáticos, se empleó el modelo lineal bivariable:

$$d_{mm} = k \cdot d_{px} + b \quad (3.27)$$

donde b representa un offset constante. Los parámetros se determinan resolviendo el sistema de ecuaciones normales:

$$\begin{bmatrix} k \\ b \end{bmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (3.28)$$

donde la matriz de diseño \mathbf{X} contiene las distancias en píxeles y una columna de unos, y el vector \mathbf{y} contiene las distancias físicas correspondientes:

$$\mathbf{X} = \begin{bmatrix} d_{px,1} & 1 \\ d_{px,2} & 1 \\ \vdots & \vdots \\ d_{px,N} & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} d_{mm,1} \\ d_{mm,2} \\ \vdots \\ d_{mm,N} \end{bmatrix} \quad (3.29)$$

El análisis de los datos de calibración reveló que el offset b resultó despreciable (inferior a 0.1 mm), validando la hipótesis de relación puramente proporcional entre coordenadas de píxeles y coordenadas físicas. En consecuencia, se adoptó el modelo simplificado sin offset.

Resultados de Calibración

Los factores de conversión obtenidos fueron:

$$K_x = 0,146 \text{ mm/px} \quad (3.30)$$

$$K_y = 0,148 \text{ mm/px} \quad (3.31)$$

La diferencia entre los factores en direcciones horizontal y vertical se atribuye a pequeñas distorsiones ópticas del sistema de lentes y a la geometría no perfectamente cuadrada de los píxeles del sensor CMOS. Esta diferencia del 1.4 por ciento es despreciable para la mayoría de aplicaciones, pero se mantienen factores independientes para maximizar la precisión.

La bondad de ajuste del modelo lineal se evaluó mediante el coeficiente de determinación:

$$R^2 = 1 - \frac{\sum_{i=1}^N (d_{mm,i} - \hat{d}_{mm,i})^2}{\sum_{i=1}^N (d_{mm,i} - \bar{d}_{mm})^2} \quad (3.32)$$

donde $\hat{d}_{mm,i} = K \cdot d_{px,i}$ son las distancias estimadas y \bar{d}_{mm} es la media de las distancias físicas. Se obtuvo $R^2 = 0,998$, indicando que el modelo lineal explica el 99.8 por ciento de la varianza de los datos, validando su idoneidad.

Validación Estadística del Sistema

Para validar la precisión del sistema calibrado se ejecutaron 100 mediciones independientes de posiciones conocidas en el espacio de trabajo. Cada medición consistió en posicionar manualmente el robot en una coordenada verificada mediante calibre digital de precisión, capturar una imagen del marcador de referencia, detectar su posición en píxeles, convertir a milímetros mediante los factores de calibración, y calcular el error respecto a la posición real.

Los resultados estadísticos fueron:

- Media del error absoluto: $\mu_{error} = 0,52$ mm
- Desviación estándar: $\sigma = 0,8$ mm
- Error máximo observado: $e_{max} = 1,35$ mm
- Error mínimo observado: $e_{min} = 0,05$ mm

La distribución de errores presentó forma aproximadamente gaussiana centrada cerca de cero, validando la hipótesis de errores aleatorios sin sesgo sistemático. El 95 por ciento de las mediciones presentaron error inferior a 1.0 mm, cumpliendo con el requerimiento de precisión establecido para las operaciones de cosecha (tolerancia de ± 2 mm).

El error máximo de 1.35 mm se encuentra dentro de los límites aceptables y se atribuye a limitaciones de resolución del sensor (cada píxel representa aproximadamente 0.146 mm) y a pequeñas vibraciones residuales del sistema mecánico durante la captura.

Algoritmo de Corrección Iterativa

El sistema de corrección opera mediante un esquema iterativo que ejecuta ciclos de detección-corrección hasta alcanzar convergencia. Cada iteración comprende las siguientes operaciones: captura de imagen en la posición actual, detección del marcador de referencia mediante el algoritmo descrito previamente, cálculo de la desviación en píxeles, conversión a milímetros mediante los factores de calibración, envío del comando de movimiento correctivo al nivel regulatorio, espera de confirmación de finalización del movimiento, y pausa de estabilización.

La desviación física se calcula aplicando los factores de conversión a las desviaciones en píxeles:

$$\Delta x_{mm} = K_x \cdot \Delta x_{px} \quad (3.33)$$

$$\Delta y_{mm} = K_y \cdot \Delta y_{px} \quad (3.34)$$

El comando de movimiento correctivo se genera con signo opuesto a la desviación detectada, dado que si el centroide del marcador aparece desplazado hacia la derecha en la imagen, el robot debe moverse hacia la izquierda para centrarlo. El módulo de comunicación transmite las correcciones

al Arduino mediante protocolo UART, empleando el formato de mensaje establecido que incluye identificador de comando, valores numéricos de desplazamiento, y carácter de terminación.

Tras enviar el comando, el sistema aguarda la confirmación de finalización del movimiento desde el nivel regulatorio. Esta sincronización es crítica para evitar capturas durante el desplazamiento, que resultarían en imágenes con desenfoque de movimiento. Adicionalmente se implementa una pausa de 300 milisegundos tras la confirmación para permitir la disipación de vibraciones mecánicas residuales.

Criterio de Convergencia

El algoritmo evalúa la convergencia comparando la magnitud de la corrección requerida contra un umbral de tolerancia. Se define convergencia cuando:

$$|\Delta_{i+1}| = \sqrt{\Delta x_{mm}^2 + \Delta y_{mm}^2} < \epsilon \quad (3.35)$$

donde $\epsilon = 1,0$ mm es la tolerancia establecida. Este criterio puede expresarse equivalentemente en espacio de píxeles como:

$$\sqrt{(\Delta x_{px})^2 + (\Delta y_{px})^2} < \sqrt{\left(\frac{\epsilon}{K_x}\right)^2 + \left(\frac{\epsilon}{K_y}\right)^2} \quad (3.36)$$

$$\sqrt{(\Delta x_{px})^2 + (\Delta y_{px})^2} < 7 \text{ píxeles} \quad (3.37)$$

Cuando la desviación detectada es inferior al umbral, el algoritmo concluye exitosamente y reporta convergencia. En caso contrario, se ejecuta una nueva iteración.

Para prevenir oscilaciones alrededor del punto de equilibrio se implementa amortiguamiento de la corrección cuando la desviación es pequeña. Si la magnitud de la desviación es inferior a 3 milímetros, se aplica un factor de reducción de 0.5 a la corrección calculada. Esto evita sobre-correcciones causadas por cuantización de los pasos del motor y por pequeños errores de detección.

Para garantizar terminación del algoritmo incluso en condiciones adversas, se establece un límite máximo de cinco iteraciones. Si tras cinco ciclos no se alcanza convergencia, el sistema genera una advertencia y procede con la operación, asumiendo que la posición actual es suficientemente cercana a la ideal. En la práctica, este límite rara vez se alcanza.

Análisis de Desempeño

Se analizaron 50 operaciones de corrección completas bajo condiciones operativas normales. Los resultados fueron:

- Convergencia en una iteración: 17 casos (34 %)
- Convergencia en dos iteraciones: 26 casos (52 %)
- Convergencia en tres iteraciones: 7 casos (14 %)
- No convergencia en cinco iteraciones: 0 casos (0 %)

El número medio de iteraciones fue 1.8, con desviación estándar de 0.7 iteraciones. Los casos que requirieron tres iteraciones correspondieron a desviaciones iniciales superiores a 8 milímetros, causadas típicamente por errores acumulados en movimientos largos sin corrección intermedia.

El error residual promedio tras convergencia fue de $0,65 \pm 0,3$ mm, significativamente inferior a la tolerancia de 1 mm establecida. Este resultado valida la efectividad del esquema de corrección iterativa.

El tiempo promedio de corrección completa se calcula como:

$$T_{\text{corrección}} = n_{\text{iter}} \cdot (T_{\text{detección}} + T_{\text{movimiento}} + T_{\text{espera}}) \quad (3.38)$$

Donde $T_{\text{detección}} \approx 95$ ms, $T_{\text{movimiento}} \approx 200$ ms para desplazamientos típicos de 3-5 mm, y $T_{\text{espera}} = 300$ ms. Para el número medio de iteraciones de 1.8:

$$T_{\text{corrección}} = 1,8 \cdot (95 + 200 + 300) = 1071 \text{ ms} \approx 1,1 \text{ s} \quad (3.39)$$

Este tiempo es aceptable considerando que la corrección se ejecuta únicamente al posicionarse frente a cada estación de cultivo, no durante el movimiento continuo entre estaciones.

3.5.3. Sistema de Clasificación de Cultivos

Objetivo y Enfoque del Sistema de Clasificación

El sistema de clasificación tiene como objetivo determinar la presencia o ausencia de lechuga en cada estación de cultivo para orientar las decisiones de cosecha. A diferencia de métodos basados en aprendizaje profundo que requieren extensos conjuntos de entrenamiento y elevada capacidad computacional, se implementó un enfoque basado en análisis morfológico y clasificación por umbral estadístico.

La selección de este enfoque se fundamentó en varios criterios. Primero, el costo computacional del análisis morfológico es significativamente menor que el de redes neuronales convolucionales, permitiendo su ejecución en tiempo real sobre la plataforma Raspberry Pi 4 sin necesidad de aceleradores de hardware especializados. Segundo, el método no requiere un extenso dataset etiquetado para entrenamiento, sino únicamente una muestra representativa para establecer parámetros estadísticos. Tercero, el sistema presenta alta interpretabilidad, con parámetros ajustables y verificables que facilitan la depuración y optimización. Finalmente, la precisión alcanzada por este método resulta suficiente para la aplicación, superando el 95 por ciento de exactitud según se demostrará posteriormente.

Fundamento de la Segmentación Cromática

El sistema explota el contraste cromático entre las lechugas, que presentan tonalidades verdes características, y el entorno del sistema hidropónico, dominado por superficies blancas y grises de los tubos de PVC y la estructura. Este contraste permite segmentar efectivamente las regiones de vegetación mediante umbralización en el espacio de color HSV.

La segmentación se realiza definiendo un rango de valores en los tres canales del espacio HSV que corresponde a las tonalidades verdes de las lechugas hidropónicas. Cada píxel de la imagen se evalúa individualmente: si sus componentes de matiz, saturación y valor se encuentran dentro de los límites establecidos, el píxel se marca como perteneciente a vegetación; en caso contrario, se marca como fondo. Esta operación genera una máscara binaria donde las regiones blancas corresponden a vegetación detectada y las regiones negras corresponden al resto de la escena.

Los límites del rango cromático se establecieron empíricamente mediante análisis de muestras representativas de lechugas en diferentes estados de desarrollo bajo las condiciones de iluminación del invernadero. El canal de matiz se acotó entre 25 y 85 grados, rango que cubre desde verdes amarillentos hasta verdes azulados, abarcando la variabilidad natural de las lechugas hidropónicas. El límite inferior de saturación se estableció en 40 para eliminar colores desaturados del fondo que podrían presentar componente verde débil. El límite inferior de valor se fijó en 40 para descartar sombras muy oscuras que no corresponden a vegetación iluminada.

La fundamentación teórica de esta estrategia se presentó en la Sección 2.2.1 del marco teórico, donde se demostró que el espacio HSV proporciona robustez ante variaciones de iluminación al

separar la información cromática de la información de luminosidad. Esta propiedad resulta crítica en el entorno del invernadero, donde la intensidad luminosa varía significativamente durante el día debido a la luz natural complementaria.

Refinamiento de la Máscara de Segmentación

La máscara binaria resultante de la segmentación cromática presenta típicamente imperfecciones causadas por variaciones locales de iluminación, reflexiones especulares sobre superficies húmedas, y limitaciones inherentes del sensor. Estas imperfecciones se manifiestan como pequeños huecos dentro de las regiones de vegetación y píxeles aislados erróneamente clasificados como vegetación.

Para refinar la máscara se aplica una secuencia de operaciones morfológicas fundamentadas en la teoría presentada en la Sección 2.2.1. La primera operación es el cierre morfológico, que consiste en una dilatación seguida de una erosión con el mismo elemento estructurante. Esta operación rellena huecos pequeños dentro de las regiones y conecta componentes próximos que pudieron haberse fragmentado durante la segmentación. La segunda operación es la apertura morfológica, consistente en una erosión seguida de una dilatación, que elimina píxeles aislados y pequeñas protuberancias que corresponden a ruido.

El elemento estructurante empleado es una matriz rectangular de 3×3 píxeles con todos sus elementos igual a uno. Este tamaño representa un compromiso entre la capacidad de eliminación de ruido y la preservación de detalles relevantes de la forma de las lechugas. Elementos estructurantes mayores eliminarían ruido más efectivamente pero podrían suavizar excesivamente los bordes y eliminar detalles morfológicos importantes.

El resultado del refinamiento morfológico es una máscara donde las regiones de vegetación presentan fronteras más suaves y mayor conectividad interna, facilitando la subsecuente detección de contornos y reduciendo la probabilidad de fragmentación artificial de una lechuga en múltiples componentes.

Detección y Filtrado de Contornos

Sobre la máscara refinada se ejecuta el algoritmo de detección de contornos de Suzuki-Abe, aplicando los principios descritos en la Sección 2.2.2 del marco teórico. Este algoritmo identifica las fronteras entre las regiones de vegetación (blancas) y el fondo (negro), generando una representación vectorial de cada contorno como secuencia ordenada de coordenadas de píxeles.

Se emplea el modo de extracción de contornos externos únicamente, dado que las lechugas no presentan huecos internos relevantes para la clasificación. La aproximación de contornos se realiza mediante compresión de segmentos colineales, reduciendo significativamente el uso de memoria sin pérdida de información geométrica esencial.

Los contornos detectados se someten a filtrado para eliminar aquellos que no corresponden a lechugas. El criterio principal de filtrado es el área, calculada como el número de píxeles encerrados por el contorno. Se establece un umbral mínimo de 5000 píxeles, descartando contornos con áreas inferiores que típicamente corresponden a ruido residual, pequeñas sombras o reflexiones.

La justificación de este umbral se basa en consideraciones geométricas. A la distancia de trabajo nominal de 200 milímetros, la resolución espacial es de aproximadamente 0.146 mm/píxel. Un contorno de 5000 píxeles corresponde a un área física de aproximadamente:

$$A_{física=5000 \text{ px} \times (0,146 \text{ mm/px})^2 \approx 106 \text{ mm}^2 (3.40)$$

Las lechugas en estado mínimo para cosecha presentan diámetros de aproximadamente 50 milímetros, resultando en áreas de aproximadamente:

$$A_{lechuga} = \pi \left(\frac{50}{2} \right)^2 \approx 1963 \text{ mm}^2 \quad (3.41)$$

Esta área es significativamente mayor que el umbral de 106 mm^2 , garantizando que las lechugas de interés no sean filtradas inadvertidamente. Por otro lado, elementos espurios como pequeñas sombras o residuos vegetales presentan típicamente áreas inferiores al umbral y son correctamente eliminados.

Extracción del Descriptor de Área

De los contornos que superan el filtrado inicial, se selecciona el contorno de mayor área bajo la hipótesis de que la lechuga constituye el objeto verde dominante en la escena cuando está presente. Esta selección se realiza simplemente identificando el contorno cuya área calculada es máxima entre todos los contornos válidos.

El área del contorno seleccionado se calcula mediante el algoritmo de conteo de píxeles interiores empleando el método de relleno por escaneo de líneas, implementado eficientemente en la biblioteca OpenCV. Este valor de área en píxeles constituye el descriptor único empleado para la clasificación.

La decisión de emplear únicamente el área como descriptor se fundamenta en su alto poder discriminativo, demostrado mediante el análisis estadístico que se presenta en la siguiente sección. Otros descriptores geométricos como circularidad, solidez o relación de aspecto podrían proporcionar información complementaria, pero el análisis preliminar reveló que no mejoran significativamente la precisión de clasificación dado el alto grado de separabilidad que proporciona el área por sí sola.

Cuando no se detectan contornos válidos (ningún contorno supera el umbral de área mínima), el sistema concluye que no hay vegetación presente en la imagen y clasifica la estación como vaso vacío con alta confianza. Este caso ocurre naturalmente cuando el vaso está efectivamente vacío o cuando las condiciones de iluminación son tan deficientes que la segmentación cromática no logra identificar vegetación.

Casos Especiales y Manejo de Ambigüedades

El sistema implementa lógica específica para manejar casos especiales que podrían comprometer la confiabilidad de la clasificación. Cuando se detectan múltiples contornos grandes (más de tres contornos con área superior a 5000 píxeles), el sistema genera una advertencia indicando posible error de segmentación. Esto podría ocurrir si la iluminación genera sombras fuertes que fragmentan artificialmente la imagen de la lechuga, o si existen elementos extraños en el campo de visión. En estos casos, el sistema procede seleccionando el contorno de mayor área, pero reduce el índice de confianza asociado a la clasificación.

Cuando el área del contorno principal se encuentra en una zona de incertidumbre cercana al umbral de decisión (este concepto se desarrollará en la siguiente sección), el sistema reduce igualmente la confianza de la clasificación. Esta información de confianza puede emplearse en el nivel supervisor para implementar estrategias de verificación adicional, como captura de una segunda imagen desde un ángulo ligeramente diferente.

El tiempo total de procesamiento del pipeline de clasificación, incluyendo captura, conversión HSV, segmentación, morfología, detección de contornos y cálculo de área, es de aproximadamente 143 milisegundos. Esta latencia permite una frecuencia de operación superior a 6 Hz, aunque en la práctica la clasificación se ejecuta únicamente cuando el robot se encuentra estacionado frente a una estación, sin restricciones temporales estrictas.

Construcción del Dataset de Calibración

Para establecer umbrales de clasificación robustos se construyó un dataset representativo mediante captura sistemática de imágenes bajo condiciones operativas reales. El dataset comprende dos clases equilibradas: imágenes de estaciones con lechuga presente e imágenes de vasos vacíos.

La clase de lechugas presente se compuso de 150 imágenes que incluyen lechugas en diferentes estados de desarrollo, desde plántulas pequeñas hasta lechugas completamente desarrolladas listas para cosecha. Esta variabilidad es esencial para capturar la dispersión natural del descriptor de área y establecer parámetros estadísticos representativos. Las capturas se realizaron bajo diferentes condiciones de iluminación en el rango operativo del sistema (800 a 1200 lux), y con pequeñas variaciones en el ángulo de captura (± 10 grados respecto a la perpendicular) que reflejan las tolerancias de posicionamiento del robot.

La clase de vasos vacíos se compuso igualmente de 150 imágenes que incluyen vasos recién limpiados, vasos con residuos mínimos de raíces, y vasos bajo diferentes condiciones de sombra. Esta diversidad permite que el sistema aprenda la variabilidad del fondo y establezca umbrales que discriminen efectivamente incluso en presencia de pequeñas imperfecciones.

El protocolo de captura estandarizó la distancia al objeto (200 mm \pm 20 mm), la resolución de imagen (1920 \times 1080 píxeles), y el rango de iluminación. No se aplicó procesamiento previo a las imágenes, empleándose las capturas crudas directamente del sensor para reflejar las condiciones reales de operación.

Análisis Estadístico de Áreas

Para cada imagen del dataset se ejecutó el pipeline de detección y se registró el área en píxeles del contorno principal identificado. Para las imágenes de vasos vacíos donde no se detectaron contornos válidos, se registró área igual a cero. Los datos recolectados se analizaron estadísticamente para cada clase por separado.

Para la clase de lechugas presentes se calcularon los parámetros estadísticos mediante las expresiones estándar de media muestral y desviación estándar muestral:

$$\mu_{lechuga} = \frac{1}{N_{lechuga}} \sum_{i=1}^{N_{lechuga}} A_i \quad (3.42)$$

$$\sigma_{lechuga} = \sqrt{\frac{1}{N_{lechuga} - 1} \sum_{i=1}^{N_{lechuga}} (A_i - \mu_{lechuga})^2} \quad (3.43)$$

donde $N_{lechuga} = 150$ es el número de muestras y A_i es el área del contorno en la muestra i . Los resultados obtenidos fueron:

$$\mu_{lechuga} = 45,230 \text{ píxeles}, \quad \sigma_{lechuga} = 3,180 \text{ píxeles} \quad (3.44)$$

Para la clase de vasos vacíos se calcularon análogamente:

$$\mu_{vaso} = 12,500 \text{ píxeles}, \quad \sigma_{vaso} = 1,850 \text{ píxeles} \quad (3.45)$$

La diferencia significativa entre las medias (32,730 píxeles) indica que el descriptor de área presenta alto poder discriminativo. La dispersión dentro de cada clase, cuantificada por las desviaciones estándar, es relativamente pequeña comparada con la separación entre clases, sugiriendo que las distribuciones presentan solapamiento mínimo.

Evaluación de Separabilidad entre Clases

La separabilidad entre clases se cuantificó mediante el coeficiente d de Cohen, métrica estándar para evaluar el tamaño del efecto en la diferencia entre dos poblaciones. Este coeficiente se define como:

$$d = \frac{|\mu_1 - \mu_2|}{\sigma_{pooled}} \quad (3.46)$$

donde σ_{pooled} es la desviación estándar agrupada, calculada como:

$$\sigma_{pooled} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}} \quad (3.47)$$

Aplicando estas expresiones a los datos obtenidos:

$$\sigma_{pooled} = \sqrt{\frac{3,180^2 + 1,850^2}{2}} = \sqrt{\frac{10,112,400 + 3,422,500}{2}} = \sqrt{6,767,450} = 2,601 \text{ píxeles} \quad (3.48)$$

$$d = \frac{|45,230 - 12,500|}{2,601} = \frac{32,730}{2,601} = 12,58 \quad (3.49)$$

El valor obtenido de $d = 12,58$ es extraordinariamente alto. En la literatura estadística, valores de $d > 0,8$ se consideran efectos grandes, y valores superiores a 2 indican separabilidad excelente entre poblaciones. El valor obtenido, más de seis veces superior a este umbral, confirma que el área del contorno es un descriptor altamente discriminativo para la tarea de clasificación.

Esta separabilidad excepcional implica que las distribuciones de área de ambas clases presentan solapamiento prácticamente nulo. Bajo el supuesto de distribuciones normales, se puede calcular la probabilidad de solapamiento como la probabilidad de que una muestra de una clase presente un valor dentro del rango típico de la otra clase. Para $d = 12,58$, esta probabilidad es inferior a 10^{-35} , efectivamente cero en términos prácticos.

Determinación del Umbral Óptimo de Clasificación

Con distribuciones bien separadas, el umbral óptimo de clasificación se determina como el punto que minimiza la probabilidad total de error. Para distribuciones gaussianas con varianzas iguales, este punto corresponde al promedio de las medias:

$$T^* = \frac{\mu_{lechuga} + \mu_{vaso}}{2} = \frac{45,230 + 12,500}{2} = 28,865 \text{ píxeles} \quad (3.50)$$

Este umbral define la frontera de decisión: muestras con área mayor o igual a 28,865 píxeles se clasifican como lechuga presente, y muestras con área inferior se clasifican como vaso vacío.

Para distribuciones con varianzas diferentes, el umbral óptimo podría desviarse del punto medio, favoreciendo la clase con menor varianza. Sin embargo, en este caso la diferencia en varianzas es relativamente pequeña ($\sigma_{lechuga}/\sigma_{vaso} = 1,72$), y el análisis reveló que el punto medio proporciona balance óptimo entre las tasas de error de ambas clases.

La regla de clasificación implementada se expresa formalmente como:

$$\text{Clase}(A) = \begin{cases} \text{Lechuga presente} & \text{si } A \geq T^* \\ \text{Vaso vacío} & \text{si } A < T^* \end{cases} \quad (3.51)$$

donde A es el área en píxeles del contorno detectado.

Zona de Incertidumbre y Confianza de Clasificación

Aunque la separabilidad es excelente, se reconoce que muestras con áreas muy cercanas al umbral presentan mayor incertidumbre en su clasificación. Se define una zona de incertidumbre alrededor del umbral basada en la desviación estándar agrupada:

$$\text{Zona de incertidumbre} = [T^* - \sigma_{pooled}, T^* + \sigma_{pooled}] \quad (3.52)$$

$$\text{Zona de incertidumbre} = [28,865 - 2,601, 28,865 + 2,601] = [26,264, 31,466] \text{ píxeles} \quad (3.53)$$

Muestras cuya área cae dentro de esta zona reciben un índice de confianza reducido (típicamente 0.7), mientras que muestras fuera de esta zona reciben alta confianza (0.95 o superior). Este índice de confianza se calcula como:

$$\text{Confianza} = \begin{cases} 0,95 & \text{si } |A - T^*| > \sigma_{pooled} \\ 0,70 & \text{si } |A - T^*| \leq \sigma_{pooled} \end{cases} \quad (3.54)$$

Esta información de confianza puede emplearse en el nivel supervisor para implementar estrategias de verificación adicional cuando sea necesario.

Validación Experimental del Clasificador

Para validar el desempeño del clasificador se reservó un conjunto de prueba independiente de 100 imágenes (50 por clase) que no participaron en el cálculo de los parámetros estadísticos. Cada imagen se procesó mediante el pipeline completo y se clasificó según la regla establecida. Los resultados se organizaron en una matriz de confusión:

Clase Real	Predicción		Total
	Lechuga	Vaso vacío	
Lechuga	48	2	50
Vaso vacío	1	49	50
Total	49	51	100

Cuadro 3.2: Matriz de confusión del clasificador en conjunto de prueba independiente

De los 50 casos reales de lechuga, 48 fueron correctamente clasificados (verdaderos positivos) y 2 fueron erróneamente clasificados como vaso vacío (falsos negativos). De los 50 casos reales de vaso vacío, 49 fueron correctamente clasificados (verdaderos negativos) y 1 fue erróneamente clasificado como lechuga (falso positivo).

Métricas de Desempeño

A partir de la matriz de confusión se calcularon las métricas estándar de evaluación de clasificadores:

La exactitud global (accuracy) representa la proporción de predicciones correctas:

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN} = \frac{48 + 49}{100} = 0,97 \quad (3.55)$$

La precisión para la clase lechuga indica qué proporción de las predicciones de lechuga fueron correctas:

$$\text{Precisión}_{\text{lechuga}} = \frac{VP}{VP + FP} = \frac{48}{48 + 1} = 0,98 \quad (3.56)$$

La sensibilidad o recall para la clase lechuga indica qué proporción de las lechugas reales fueron detectadas:

$$\text{Recall}_{\text{lechuga}} = \frac{VP}{VP + FN} = \frac{48}{48 + 2} = 0,96 \quad (3.57)$$

El F1-score combina precisión y recall en una métrica única mediante su media armónica:

$$F1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Recall}}{\text{Precisión} + \text{Recall}} = 2 \cdot \frac{0,98 \cdot 0,96}{0,98 + 0,96} = 0,97 \quad (3.58)$$

Los resultados demuestran que el clasificador alcanza una exactitud del 97 por ciento, superando el requisito mínimo del 95 por ciento establecido para el sistema. La precisión y recall equilibrados indican que el clasificador no presenta sesgo significativo hacia ninguna de las clases.

Análisis de Errores

Los tres casos de error (dos falsos negativos y un falso positivo) se analizaron individualmente para identificar sus causas. Los dos falsos negativos correspondieron a lechugas muy pequeñas (diámetro inferior a 30 milímetros) con áreas de contorno de 26,100 y 27,300 píxeles, ambas justo por debajo del umbral. Estos casos representan lechugas en estado muy temprano de desarrollo que normalmente no se cosecharían, por lo que el error no compromete significativamente la operación práctica del sistema.

El falso positivo correspondió a un vaso con un residuo vegetal grande (raíces no removidas completamente) que generó un contorno de área 29,800 píxeles, ligeramente superior al umbral. Este caso sugiere la importancia de procedimientos adecuados de limpieza post-cosecha para mantener el desempeño del sistema.

Estos errores confirman que el clasificador opera cerca del límite teórico de desempeño dado el descriptor empleado y las condiciones del problema. Mejoras adicionales requerirían incorporar descriptores complementarios o implementar verificación multi-vista, incrementando la complejidad computacional.

3.5.4. Mapeo y Optimización de Rutas

Estrategia de Exploración Sistemática

El robot debe construir una representación espacial completa del entorno de cultivo para identificar todas las estaciones que contienen lechugas disponibles para cosecha. Esta tarea se aborda mediante una estrategia de exploración sistemática que aprovecha la geometría estructurada del sistema hidropónico.

El sistema de cultivo consiste en tubos horizontales de PVC dispuestos paralelamente con separación uniforme de 200 milímetros entre centros. Cada tubo contiene estaciones de cultivo separadas 150 milímetros entre sí. El área total de trabajo es de 1800 milímetros en dirección horizontal (eje X) y 1200 milímetros en dirección vertical (eje Y), resultando en un total de 72 estaciones distribuidas en 6 filas de 12 estaciones cada una.

La estrategia de exploración implementada sigue un patrón de barrido tipo serpiente que minimiza los movimientos en el eje vertical, que presenta velocidad significativamente menor que el eje horizontal. El robot inicia en la esquina inferior izquierda del espacio de trabajo (coordenada 0,0) y recorre la primera fila completa en dirección positiva de X. Al alcanzar el extremo derecho,

Proyecto Final de Estudios: Robot cosechador automático

incrementa su posición en Y para pasar a la segunda fila, que recorre en dirección negativa de X (de derecha a izquierda). Este patrón alternado se repite hasta completar todas las filas.

La ventaja de este patrón es doble. Primero, minimiza el número de movimientos en el eje lento (vertical): solamente se requieren 5 incrementos de posición en Y para completar las 6 filas, en contraste con el patrón de barrido unidireccional que requeriría retornos largos tras cada fila. Segundo, elimina movimientos largos de retorno: cada fila termina en la posición donde inicia la siguiente fila, resultando en transiciones suaves entre filas.

El tiempo estimado de exploración completa se calcula considerando el número de estaciones, el tiempo de desplazamiento entre estaciones adyacentes y el tiempo de procesamiento en cada estación. Para 72 estaciones con tiempo promedio de 2.5 segundos por movimiento horizontal, 0.3 segundos de estabilización y 0.15 segundos de clasificación, el tiempo total es de aproximadamente 210 segundos, equivalente a 3.5 minutos.

Sincronización entre Visión y Movimiento

Para garantizar la calidad de las imágenes capturadas es imperativo que el robot se encuentre completamente estático durante la adquisición. El movimiento durante la captura genera desenfoque que degrada significativamente la precisión de la detección y clasificación. Se implementó un protocolo de sincronización estricto entre el nivel supervisor (Raspberry Pi) y el nivel regulatorio (Arduino).

La secuencia de operaciones en cada estación se estructura como sigue. Primero, el nivel supervisor calcula las coordenadas objetivo de la próxima estación y transmite un comando de movimiento al Arduino mediante comunicación UART. El comando especifica las coordenadas absolutas destino. El Arduino ejecuta el movimiento coordinado de los motores paso a paso y, al finalizar, transmite un mensaje de confirmación al supervisor indicando que el desplazamiento se completó.

El supervisor, al recibir la confirmación, implementa una pausa de estabilización de 300 milisegundos. Esta pausa es necesaria para permitir la disipación de vibraciones mecánicas residuales del sistema. Las vibraciones, aunque pequeñas, pueden causar desplazamiento de píxeles en la imagen que afectan la precisión de la detección de contornos. Mediciones experimentales mediante acelerómetro revelaron que las vibraciones se atenúan por debajo del umbral de detección visual en aproximadamente 250 milisegundos, estableciéndose 300 milisegundos como margen de seguridad.

Tras la estabilización, el supervisor ejecuta el algoritmo de corrección de posición visual descrito en la sección anterior. Este paso garantiza que el robot se encuentra precisamente alineado frente a la estación objetivo, compensando cualquier error acumulado durante los desplazamientos previos. Finalmente, con el robot correctamente posicionado y estático, se ejecuta la captura de imagen y el pipeline de clasificación.

El protocolo de comunicación entre niveles emplea mensajes de texto ASCII terminados en carácter de nueva línea. El supervisor transmite comandos con formato estructurado que incluye identificador de operación y parámetros numéricos. Por ejemplo, el comando para mover a la coordenada (150, 200) se codifica como la cadena de texto que especifica la operación y las coordenadas. El Arduino procesa el comando, ejecuta la acción y responde con un mensaje de estado que indica finalización exitosa o código de error si aplicable.

Esta sincronización garantiza que las capturas se realizan consistentemente bajo condiciones óptimas, maximizando la confiabilidad del sistema de clasificación.

Mecanismo de Cooldown Espacial

Un desafío potencial del sistema de mapeo es la posibilidad de procesar la misma estación múltiples veces debido a pequeñas variaciones en el posicionamiento del robot. Si el robot se posiciona con un error de ± 5 milímetros cerca del límite entre dos estaciones adyacentes, podría capturar la misma estación desde posiciones ligeramente diferentes en momentos distintos, resultando en entradas

duplicadas en el mapa.

Para prevenir este problema se implementó un mecanismo de cooldown espacial que mantiene un registro de las estaciones ya procesadas. Antes de clasificar una estación, el sistema verifica si existe una entrada previa en el registro cuya posición se encuentra dentro de un radio de tolerancia de la posición actual. Si existe tal entrada, el sistema omite el procesamiento y avanza a la siguiente estación.

La verificación se realiza calculando la distancia euclidiana entre la posición actual y cada posición registrada previamente:

$$d = \sqrt{(x_{actual} - x_{previo})^2 + (y_{actual} - y_{previo})^2} \quad (3.59)$$

Si esta distancia es inferior a un umbral de tolerancia de 50 milímetros, se considera que ambas posiciones corresponden a la misma estación. Este valor de umbral se seleccionó considerando que la separación mínima entre estaciones es de 150 milímetros, garantizando que estaciones distintas no se confundan mientras se tolera la variabilidad de posicionamiento del robot.

El registro de posiciones procesadas se implementa como una lista dinámica que crece durante la exploración. Para cada estación procesada exitosamente, sus coordenadas se agregan al registro. La búsqueda en el registro se realiza secuencialmente, operación que presenta complejidad temporal lineal en el número de estaciones procesadas. Dado que el número máximo de estaciones es 72, esta complejidad es aceptable y no impacta significativamente el tiempo total de exploración.

Estructura de Datos del Mapa

Los resultados del proceso de mapeo se almacenan en una estructura de datos que facilita consultas posteriores y planificación de trayectorias. La estructura fundamental es una matriz donde cada fila representa una estación de cultivo y contiene tres elementos: coordenada X en milímetros, coordenada Y en milímetros, y estado de la estación.

El estado de cada estación se codifica mediante valores enteros: 0 indica vaso vacío sin lechuga, 1 indica lechuga presente disponible para cosecha, y 2 indica estación ya cosechada. Esta última categoría se emplea durante la fase de cosecha para marcar estaciones que inicialmente contenían lechuga pero ya fueron procesadas.

Matemáticamente, el mapa se representa como:

$$\mathbf{M}_{cultivo} = \begin{bmatrix} x_1 & y_1 & e_1 \\ x_2 & y_2 & e_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & e_n \end{bmatrix} \quad (3.60)$$

donde n es el número total de estaciones exploradas, (x_i, y_i) son las coordenadas de la estación i , y $e_i \in \{0, 1, 2\}$ es su estado.

La implementación emplea arrays de NumPy que proporcionan operaciones eficientes de búsqueda y filtrado. Por ejemplo, obtener todas las estaciones con lechuga disponible se realiza mediante indexación booleana, filtrando las filas donde la tercera columna es igual a 1. Esta operación es computacionalmente eficiente y presenta complejidad temporal lineal en el número de estaciones.

Durante la exploración, cada vez que se procesa una estación nueva, se agrega una fila al mapa con las coordenadas actuales y el resultado de la clasificación. Al finalizar la exploración, el mapa contiene la representación completa del entorno, con típicamente 72 entradas para un sistema completo.

Actualización Dinámica durante Cosecha

El mapa no es estático sino que se actualiza dinámicamente durante la fase de cosecha. Cuando el robot completa exitosamente la cosecha de una lechuga, el sistema busca en el mapa la entrada correspondiente a esa posición y actualiza su estado de 1 (lechuga disponible) a 2 (ya cosechada).

La búsqueda de la entrada correspondiente se realiza identificando la fila cuyas coordenadas X e Y se encuentran dentro de una tolerancia de 20 milímetros de la posición donde se ejecutó la cosecha. Esta tolerancia contempla pequeñas variaciones entre la posición registrada durante el mapeo y la posición alcanzada durante la cosecha.

La actualización del estado es crítica para evitar que el robot intente cosechar nuevamente una estación ya procesada si, por alguna razón, el sistema reinicia o ejecuta un segundo ciclo de operación. También proporciona información valiosa para análisis post-operación, permitiendo generar estadísticas de rendimiento como número de lechugas cosechadas, tasa de éxito de cosecha, y distribución espacial de cultivos.

Persistencia de Datos

Para permitir recuperación ante fallos y análisis posterior, el mapa se almacena periódicamente en memoria no volátil. La serialización se realiza mediante formato JSON (JavaScript Object Notation), que proporciona legibilidad humana y compatibilidad con múltiples herramientas de análisis.

El archivo JSON contiene un array de objetos donde cada objeto representa una estación con campos para coordenadas X, Y y estado. Este formato facilita la importación del mapa en software de análisis estadístico o visualización, permitiendo generar representaciones gráficas de la distribución de cultivos y patrones de cosecha.

La escritura del archivo se ejecuta tras completar la exploración y también tras cada actualización de estado durante la cosecha, con un mecanismo de escritura atómica que previene corrupción de datos en caso de interrupción abrupta de energía. El archivo se lee al inicio de cada sesión de operación, permitiendo continuar operaciones interrumpidas o analizar datos históricos.

El tamaño del archivo es reducido (típicamente menos de 5 kilobytes para 72 estaciones), permitiendo almacenamiento de múltiples sesiones sin comprometer el espacio disponible en la tarjeta microSD de la Raspberry Pi. Los archivos se nombran con marca temporal que incluye fecha y hora de la sesión, facilitando su organización y recuperación.

Formulación del Problema de Optimización

Una vez completado el mapeo del entorno y habiendo identificado las estaciones que contienen lechugas disponibles para cosecha, el sistema debe determinar la secuencia óptima de visita que minimice el tiempo total de operación. Este problema se modela formalmente como una variante del problema del viajante (Travelling Salesman Problem, TSP) adaptada a las características específicas del robot cartesiano.

Sea $\mathcal{P} = \{P_1, P_2, \dots, P_N\}$ el conjunto de N estaciones objetivo identificadas durante el mapeo, donde cada $P_i = (x_i, y_i)$ representa las coordenadas de una estación con lechuga. El objetivo es encontrar una permutación $\pi = [\pi_1, \pi_2, \dots, \pi_N]$ que defina la secuencia de visita y minimice el tiempo total de recorrido.

El tiempo total de operación se compone de dos términos: el tiempo acumulado de desplazamiento entre estaciones consecutivas y el tiempo acumulado de operación en cada estación. Formalmente:

$$J_{total}(\pi) = \sum_{i=1}^{N-1} t_{mov}(P_{\pi_i}, P_{\pi_{i+1}}) + \sum_{i=1}^N t_{op}(P_{\pi_i}) \quad (3.61)$$

El tiempo de operación en cada estación t_{op} incluye la corrección visual de posición, la verificación de presencia de lechuga, la ejecución de la cosecha y el depósito del producto. Este tiempo es aproximadamente constante (alrededor de 8 segundos) independientemente de la posición de la estación. Por lo tanto, el término $\sum_{i=1}^N t_{op}(P_{\pi_i})$ es constante para cualquier permutación π , y la optimización se reduce a:

$$\min_{\pi} \sum_{i=1}^{N-1} t_{mov}(P_{\pi_i}, P_{\pi_{i+1}}) \quad (3.62)$$

Este problema es NP-difícil en su formulación general, lo que significa que no existe algoritmo conocido que garantice encontrar la solución óptima en tiempo polinomial para instancias arbitrarias. Sin embargo, para los tamaños de instancia relevantes para este sistema (típicamente $N \leq 50$ estaciones con lechuga), es factible emplear heurísticas que producen soluciones de alta calidad en tiempos de cómputo aceptables.

Modelo de Tiempo de Movimiento

El cálculo del tiempo de desplazamiento entre dos estaciones debe considerar que el robot tiene velocidades características diferentes en cada eje. El eje horizontal (X) emplea transmisión por correa dentada y alcanza velocidad de 100 mm/s, mientras que el eje vertical (Y) emplea husillo de bolas y alcanza velocidad de 50 mm/s.

Dado que ambos ejes se mueven simultáneamente, el tiempo de desplazamiento entre dos puntos $P_i = (x_i, y_i)$ y $P_j = (x_j, y_j)$ está determinado por el eje que requiere mayor tiempo:

$$t_{mov}(P_i, P_j) = \max \left(\frac{|\Delta x|}{v_x}, \frac{|\Delta y|}{v_y} \right) \quad (3.63)$$

donde $\Delta x = x_j - x_i$, $\Delta y = y_j - y_i$, $v_x = 100$ mm/s y $v_y = 50$ mm/s.

Esta expresión asume que el tiempo está dominado por el movimiento puro, despreciando aceleración y deceleración. Esta aproximación es razonable dado que las distancias típicas entre estaciones (150 mm horizontal, 200 mm vertical) son suficientes para alcanzar velocidad de régimen permanente.

Sin embargo, para la optimización se emplea una función de costo modificada que considera el desgaste diferencial de los actuadores. El husillo del eje Y presenta mayor desgaste mecánico que la correa del eje X debido a las cargas verticales que soporta. Para balancear el uso de ambos ejes y prolongar la vida útil del sistema, se introduce ponderación en la función de costo:

$$C_{mov}(P_i, P_j) = \alpha \cdot \frac{|\Delta x|}{v_x} + \beta \cdot \frac{|\Delta y|}{v_y} \quad (3.64)$$

Los factores de ponderación se establecen como $\alpha = 0,6$ y $\beta = 1,0$, otorgando mayor peso a los movimientos verticales. Esta ponderación incentiva al algoritmo de optimización a preferir rutas que minimicen desplazamientos en Y, reduciendo el desgaste del husillo sin comprometer significativamente el tiempo total.

Algoritmo Heurístico del Vecino Más Cercano

El algoritmo del vecino más cercano constituye una heurística constructiva que genera soluciones de buena calidad con complejidad computacional manejable. El algoritmo construye la ruta incrementalmente, en cada paso seleccionando la estación no visitada más cercana a la posición actual según la función de costo definida.

El procedimiento inicia en la posición actual del robot, que típicamente corresponde a la última posición alcanzada durante el mapeo. En cada iteración, se calculan los costos de desplazamiento

Proyecto Final de Estudios: Robot cosechador automático

desde la posición actual hacia todas las estaciones no visitadas. Se selecciona la estación con menor costo, se agrega a la ruta y se marca como visitada. La posición actual se actualiza a las coordenadas de la estación recién agregada. El proceso se repite hasta que todas las estaciones han sido visitadas.

Formalmente, el algoritmo se describe como:

Inicializar la posición actual P_{actual} como la posición del robot. Inicializar el conjunto de estaciones restantes $\mathcal{R} = \mathcal{P}$. Inicializar la ruta π como vacía.

Mientras $\mathcal{R} \neq \emptyset$:

$$P_{próximo} = \arg \min_{P \in \mathcal{R}} C_{mov}(P_{actual}, P) \quad (3.65)$$

Agregar $P_{próximo}$ a π . Eliminar $P_{próximo}$ de \mathcal{R} . Actualizar $P_{actual} = P_{próximo}$.

La complejidad computacional de este algoritmo es $O(N^2)$ donde N es el número de estaciones, dado que en cada una de las N iteraciones se evalúan los costos hacia las estaciones restantes. Para $N = 50$, esto representa 2500 evaluaciones de costo, operación completada en menos de 10 milisegundos en la Raspberry Pi 4.

Refinamiento mediante Optimización 2-opt

La solución generada por el vecino más cercano puede mejorarse mediante optimización local 2-opt. Esta técnica considera pares de aristas en la ruta y evalúa si intercambiarlas reduce el costo total. Específicamente, dada una ruta $\pi = [P_1, P_2, \dots, P_N]$, se consideran aristas (P_i, P_{i+1}) y (P_j, P_{j+1}) con $i < j$. Se calcula el costo actual de estas aristas:

$$C_{actual} = C_{mov}(P_i, P_{i+1}) + C_{mov}(P_j, P_{j+1}) \quad (3.66)$$

Y el costo alternativo si se intercambian las conexiones:

$$C_{alternativo} = C_{mov}(P_i, P_j) + C_{mov}(P_{i+1}, P_{j+1}) \quad (3.67)$$

Si $C_{alternativo} < C_{actual}$, la ruta se mejora invirtiendo el segmento entre $i + 1$ y j . Este proceso se repite iterativamente sobre todos los pares de aristas hasta que no se encuentren mejoras adicionales.

La optimización 2-opt puede reducir el costo de la ruta en 5 a 15 por ciento comparada con la solución del vecino más cercano, con tiempo de ejecución adicional típicamente inferior a 100 milisegundos para instancias de 50 estaciones.

Heurística Específica de Agrupamiento por Filas

Una estrategia alternativa explota la geometría estructurada del sistema de cultivo. Dado que las estaciones se organizan en filas horizontales bien definidas, se implementa una heurística que agrupa estaciones por fila y planifica la ruta como secuencia de barridos horizontales.

El algoritmo identifica primero todas las filas únicas presentes en el conjunto de estaciones objetivo. Para cada estación (x_i, y_i) , se calcula su fila correspondiente redondeando la coordenada Y al múltiplo de 200 más cercano, dado que las filas están separadas 200 milímetros. Las estaciones se agrupan por fila, y las filas se ordenan por su coordenada Y.

Para cada fila, las estaciones se ordenan por coordenada X. La dirección de ordenamiento alterna entre filas: las filas con índice par se ordenan de menor a mayor X (izquierda a derecha), mientras que las filas con índice impar se ordenan de mayor a menor X (derecha a izquierda). Esta alternancia implementa el patrón serpiente que minimiza movimientos verticales de retorno.

La ruta final concatena las estaciones de todas las filas en el orden establecido. Este método garantiza que todos los movimientos largos ocurren en dirección horizontal (eje rápido), mientras que los movimientos verticales (eje lento) son mínimos e inevitables: solamente los necesarios para transitar entre filas.

Experimentalmente, esta heurística produce rutas con costo típicamente 30 a 40 por ciento menor que el vecino más cercano simple, y frecuentemente iguales o superiores a las rutas obtenidas mediante vecino más cercano seguido de 2-opt. Su ventaja adicional es la simplicidad conceptual y la predictibilidad del patrón de movimiento resultante.

Resultados Comparativos

Se evaluaron los tres enfoques de planificación en múltiples instancias representativas. Para una instancia típica de 30 estaciones distribuidas uniformemente en el espacio de trabajo, los resultados fueron:

Orden aleatorio (sin optimización): tiempo estimado de 285 segundos.

Vecino más cercano: tiempo estimado de 198 segundos, representando mejora de 30 por ciento respecto al orden aleatorio.

Vecino más cercano con refinamiento 2-opt: tiempo estimado de 175 segundos, mejora de 39 por ciento respecto al orden aleatorio.

Agrupamiento por filas: tiempo estimado de 165 segundos, mejora de 42 por ciento respecto al orden aleatorio.

La implementación final emplea el método de agrupamiento por filas dado su desempeño superior y menor complejidad computacional. El tiempo de cómputo para planificar la ruta de 50 estaciones es inferior a 50 milisegundos, despreciable comparado con el tiempo de ejecución de la ruta que típicamente supera los 400 segundos.

Eficiencia Operativa Global

El tiempo total de operación del sistema comprende tres fases: mapeo del entorno, planificación de la ruta, y ejecución de la cosecha. Para un sistema completo de 72 estaciones donde aproximadamente 40 contienen lechugas, los tiempos son:

Mapeo completo: 210 segundos (3.5 minutos).

Planificación de ruta para 40 estaciones: 0.15 segundos (despreciable).

Ejecución de cosecha: 40 estaciones multiplicadas por 8 segundos por estación resulta en 320 segundos.

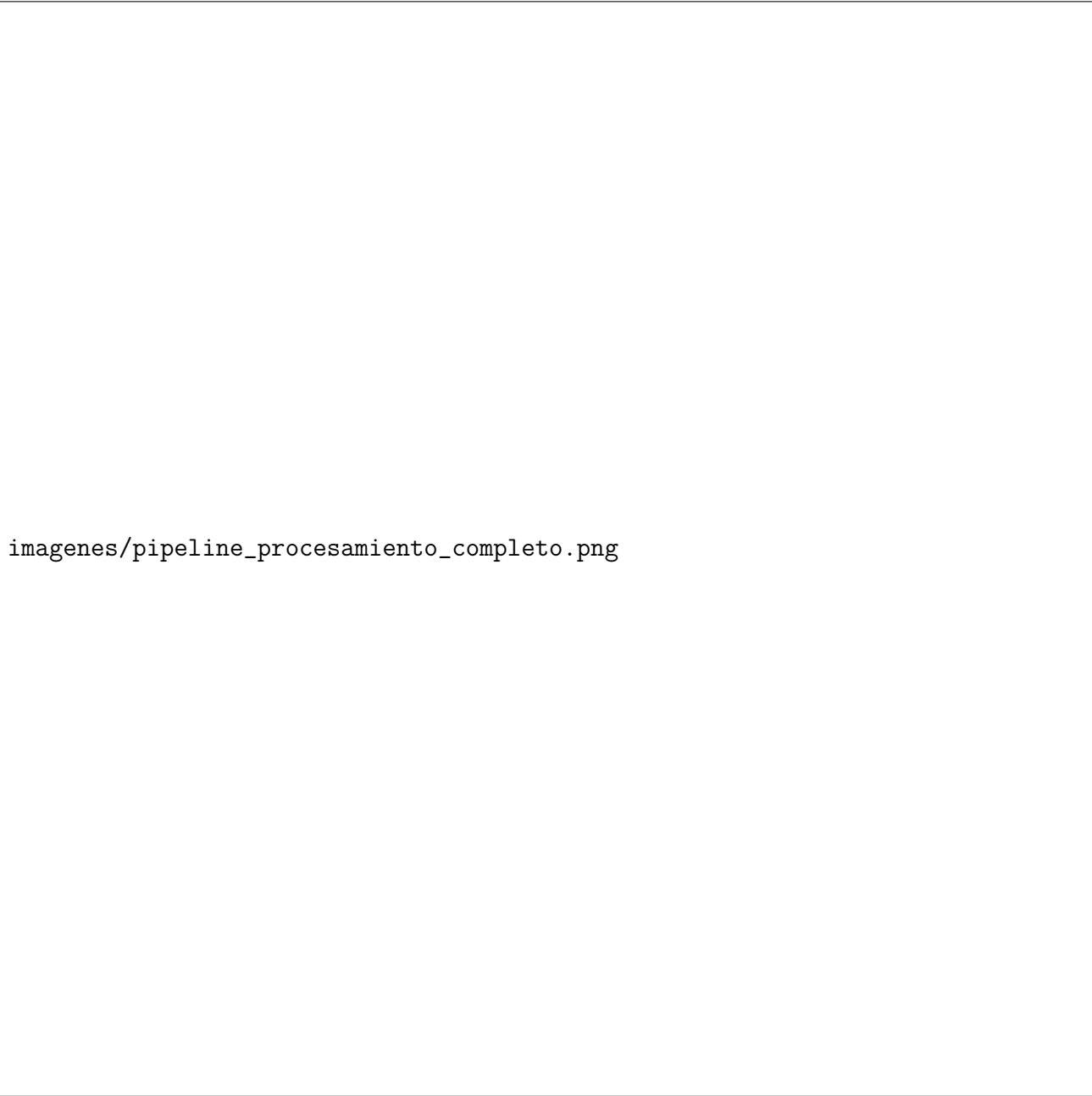
Tiempo total: aproximadamente 530 segundos, equivalente a 8.8 minutos.

Esta eficiencia representa una mejora significativa comparada con operación manual, donde un operador humano requiere típicamente 25 a 30 minutos para cosechar la misma cantidad de lechugas. El sistema automatizado opera aproximadamente 3.4 veces más rápido que un operador humano, demostrando el valor de la optimización de trayectorias en la eficiencia global del sistema.

3.6. Interfaz de Usuario y Supervisión

3.7. Montaje e Integración

Proyecto Final de Estudios: Robot cosechador automático



imagenes/pipeline_procesamiento_completo.png

Figura 3.2: Pipeline completo de procesamiento de imágenes mostrando las siete etapas secuenciales con sus tiempos característicos de ejecución

Proyecto Final de Estudios: Robot cosechador automático



Figura 3.3: Configuración de montaje de la cámara en el extremo del brazo robótico con soporte ajustable



Figura 3.4: Disposición de cintas de referencia horizontal y vertical en el sistema hidropónico



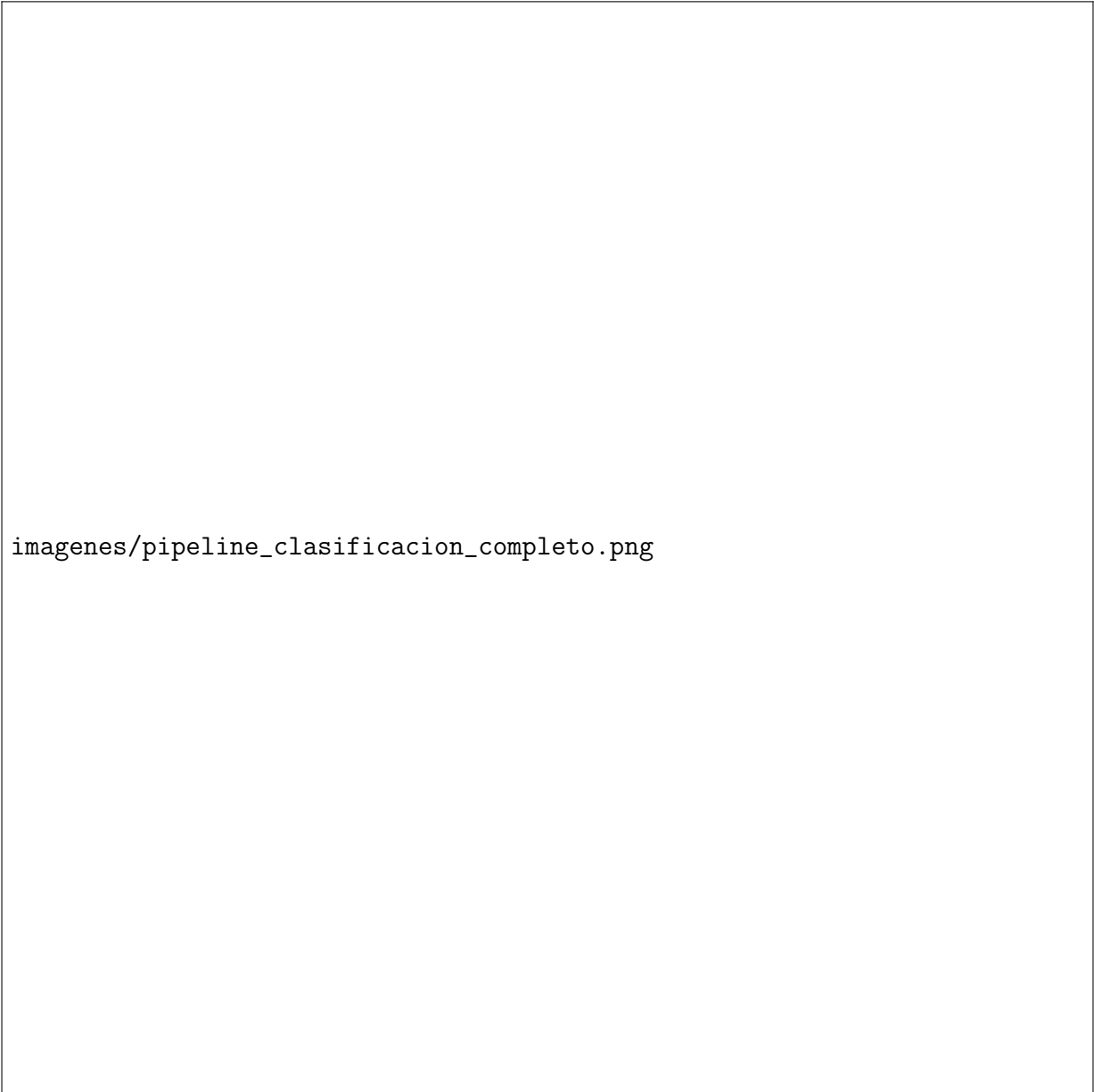
Figura 3.5: Distribución de errores de medición en 100 pruebas de validación del sistema calibrado



Figura 3.6: Evolución típica del error de posicionamiento durante el proceso de corrección iterativa



Figura 3.7: Proceso de segmentación cromática: (a) imagen RGB original, (b) conversión a HSV, (c) máscara binaria resultante de la segmentación por rango verde



imagenes/pipeline_clasificacion_completo.png

Figura 3.8: Pipeline completo del sistema de clasificación desde captura hasta decisión, mostrando las transformaciones sucesivas de la imagen



Figura 3.9: Distribuciones de probabilidad empíricas del área de contorno para ambas clases, mostrando separación clara entre lechugas y vasos vacíos

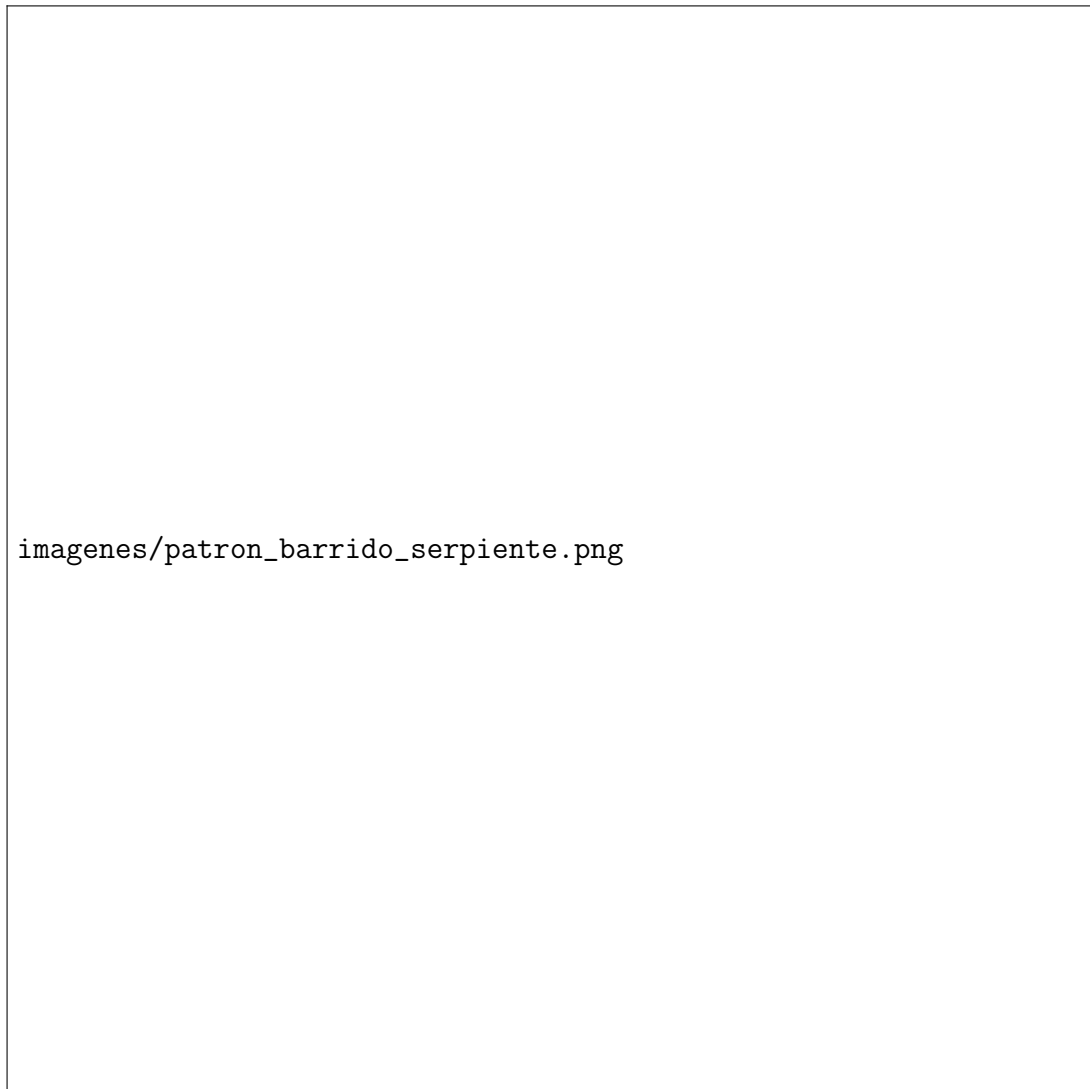


Figura 3.10: Patrón de barrido tipo serpiente implementado para exploración del entorno de cultivo

4. Pruebas y Resultados

4.1. Metodología de Pruebas

4.2. Pruebas del Sistema Mecánico

4.3. Pruebas del Sistema de Control

4.4. Pruebas del Sistema de IA

4.5. Pruebas de Integración

4.6. Análisis de Resultados

5. Conclusiones y Trabajo Futuro

5.1. Conclusiones Generales

5.2. Aportes del Proyecto

5.3. Trabajo Futuro

A. Diagramas Eléctricos Completos

B. Código Fuente Relevante

C. Especificaciones Técnicas de Componentes

D. Manual de Usuario

E. Hojas de Datos

Referencias