

Evaluación del estado de maduración de la piña en su variedad perolera mediante técnicas de visión artificial

Luis Ángel Silva

PhD en Automática y Robótica,
Universidad Politécnica de Madrid, España
Docente Tiempo Completo, Investigador Grupo de Control Industrial, Universidad Pontificia Bolivariana
Bucaramanga, Colombia
luis.angel@upb.edu.co

Sergio Lizcano

Ing. Electrónico
Universidad Pontificia Bolivariana
Investigador Grupo de Control Industrial,
Universidad Pontificia Bolivariana
Bucaramanga, Colombia
sergio.lizcano@upb.edu.co

Resumen— Los sistemas de visión artificial permiten identificar características físicas y defectos de los productos de manera no intrusiva y confiable. Debido a estas ventajas, los sistemas de visión han tenido gran aceptación en las industrias agrícola y alimenticia, ya que estas industrias requieren de una alta demanda de objetividad, coherencia y eficiencia en el control de la calidad de los productos, condiciones que los sistemas de visión artificial pueden cumplir. Este trabajo propone un procedimiento para evaluar de manera automática el estado de maduración de la piña (Ananas Comosus) variedad perolera en post-cosecha mediante técnicas de visión artificial. El procedimiento de evaluación propuesto es implementado a través de un algoritmo de procesamiento de imágenes digitales en color basado en las etapas de pre-procesamiento, segmentación, extracción de características y clasificación estadística. Se empleó, imágenes en el espacio de color HSV, segmentación por umbralización automática mediante el método de Otsu, el momento de primer orden de las distribuciones de los planos H y S como características, y el algoritmo MBSAS (Modified Basic Sequential Algorithmic Scheme— en inglés) para la clasificación. Se utilizaron 319 imágenes, de las cuales 110 imágenes fueron utilizadas en el proceso de entrenamiento y 209 imágenes fueron utilizadas en el proceso de evaluación. Los resultados obtenidos por el procedimiento de evaluación propuesto en este trabajo, fueron contrastados con el juicio de valor de tres expertos, que mostraron que el algoritmo propuesto tiene una eficiencia de evaluación cercana al 96,36%.

Palabras clave— Sistema de visión artificial, Segmentación, Extracción de características, Clasificación

Abstract— Computer vision systems allow identifying the physical characteristics and defects of a product in a non-invasive and reliable form. Due to these advantages, computer vision systems have gained acceptance in the food industry, since this industry requires a high demand for objectivity, consistency and efficiency in controlling the quality of the product, conditions that this systems can meet. This paper proposes a method for automatically evaluating the maturation state of the pineapple (Ananas Comosus) perolera variety in post-harvest using computer vision system. The proposed evaluation

procedure is implemented through an algorithm of color digital image processing based on the stages of preprocessing, segmentation, feature extraction and statistical classification. Algorithm used, images in the HSV color space, automatic segmentation using Otsu's method, the first moment of the distributions of H and S planes as features, and the algorithm MBSAS (Modified Basic Sequential Algorithmic Scheme) for classification. 319 images were used, of which 110 images were used in the process of training and 209 images were used in the evaluation process. The results of the evaluation procedure proposed in this paper were compared with the value judgment of three experts, showing that the proposed algorithm has an efficiency of close to 96.36% assessment.

Keywords— Computer vision system, segmentation, feature extraction, statistical classification.

I. INTRODUCCIÓN

La globalización de la economía y la firma de tratados de libre comercio entre los diferentes países, han traído para el sector industrial de cada país, retos en temas de sostenibilidad y competitividad a nivel internacional. Estos retos implican, un mejoramiento de los procesos industriales que garanticen un incremento en aspectos relacionados con el volumen de producción, la estandarización y el control de calidad de los productos. En los últimos años, Colombia no ha sido ajena a la firma de estos tratados y en la actualidad tiene vigentes 7 tratados de libre comercio con 15 países, entre ellos, Estados Unidos, que representa la principal economía del mundo [1].

Las condiciones geográficas de Colombia, permiten el cultivo de una extensa variedad de productos agrícolas. Esta ventaja plantea para Colombia, grandes posibilidades en el incremento de las exportaciones de productos agrícolas como

son las frutas, verduras y hortalizas [2]. Sin embargo, actualmente este sector no es plenamente competitivo en los mercados internacionales, ya que existen deficiencias en los procesos de producción industrial en términos de automatización y control de calidad de los productos. En un alto porcentaje, estos procesos se realizan de manera empírica y subjetiva por parte de un inspector de calidad, actividad que no garantiza la homogeneidad y estandarización del producto.

El control de calidad automático en productos agrícolas ha evolucionado en conjunto con la tecnología en adquisición de imágenes y la velocidad de procesamiento de los actuales computadores [3]. Actualmente, muchos países utilizan la visión artificial como una alternativa para mejorar el control de calidad en productos agrícolas, lo cual les permite ser más competitivos en el mercado globalizado pues brinda productos de mayor calidad a los consumidores.

Se puede definir la visión artificial como el área del conocimiento que estudia teorías y métodos para el análisis automático de la información contenida en imágenes. La metodología clásica de la visión artificial recorre las siguientes etapas: adquisición, pre-procesamiento y segmentación, extracción de características, reconocimiento de patrones y evaluación [4].

Una síntesis de los métodos y aplicaciones que emplean visión artificial para la evaluación de la calidad de productos alimenticios se presentan en [5]. Se hace énfasis en la evaluación de la calidad de carnes, aves, mariscos, granos, frutas y vegetales.

Revisiones bibliográficas sobre métodos no destructivos para el control de calidad de productos agrícolas son presentadas en [6] y [7]. Existen numerosos trabajos sobre la evaluación de la calidad y detección de defectos para diferentes productos agrícolas. Trabajos en manzanas son presentados en [8-10], trabajos en productos cítricos en [11-13], trabajos con fresas, aceitunas y papas en [14-18].

En Colombia, la aplicación de la visión artificial se ha centrado en trabajos con granos de café [19-23]. Estos trabajos tienen como objetivo crear un sistema de visión artificial para la selección de granos de café según la etapa de maduración.

En Santander, la piña representa el tercer producto agrícola más importante para la región, su fruto se comercializa en casi todo el país y en paí-

ses vecinos. La variedad perolera es la de mayor relevancia para la región ya que es muy apetecida por su sabor y calidad. También es muy resistente al transporte, lo que permite al agricultor tener mayor tranquilidad y confianza al momento de comercializarla fuera de la región.

En este artículo, se presenta una metodología para identificar automáticamente las etapas de maduración de la piña en su variedad perolera con técnicas de visión artificial. El sistema de visión está formado por una cámara a color y el algoritmo desarrollado permite el procesamiento de imágenes y el reconocimiento de patrones. Las imágenes son transformadas del espacio de color RGB al espacio de color HSV. La segmentación es automática y usa el método de Otsu. Como características para extraer se ha seleccionado el momento de primer orden de las distribuciones de los planos H y S. En la etapa de clasificación se ha empleado el algoritmo MBSAS [25]. Los resultados obtenidos son altamente satisfactorios y permitirán en un futuro próximo, dar a la industria de la región, una alternativa para el control de calidad de diferentes productos agrícolas de forma automática.

El resto de este artículo está estructurado de la siguiente manera. Inicialmente, se presenta la metodología empleada para determinar el grado de maduración de la piña en su variedad perolera que detalla los aspectos técnicos en cada una de las etapas del sistema de visión implementado. Posteriormente, se contrastan los resultados obtenidos mediante el algoritmo propuesto de clasificación automática con los resultados obtenidos a partir del criterio de expertos en clasificación manual. Finalmente, se presentan las conclusiones y los trabajos futuros.

II. METODOLOGÍA

La metodología propuesta emplea las técnicas de visión artificial a través de las siguientes etapas: adquisición, pre-procesamiento, segmentación, extracción de características, clasificación y evaluación.

A. Sistema de adquisición de imágenes

Las imágenes fueron tomadas mediante una cámara fotográfica digital Canon Power Shot A590 con una resolución de 1536x2048 píxeles.

Para la toma de las imágenes y con el fin de facilitar su procesamiento posteriormente, se construyó una recámara con condiciones de iluminación homogénea y controlada. La recámara fue construida en madera con dimensiones de 60cmX60cmX60cm y el interior pintado de color negro con fondo azul. El sistema de iluminación está formado por una bombilla halógena. Se hizo una ranura a través de la cual, se ajusta el lente de la cámara y se toman las imágenes. En la Fig. 1. se puede apreciar diferentes vistas de la recámara construida.

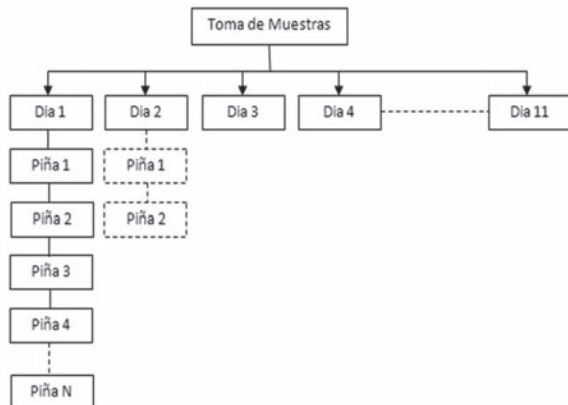
El tiempo de maduración de la piña en post-cosecha es de aproximadamente 11 días. Durante este tiempo, se tomaron fotografías diariamente a cada piña y se almacenaban en una base de datos organizada como se muestra en la Fig. 2. En total se tomaron 1500 imágenes, de las cuales, una parte fue empleada en el proceso de entrenamiento y la otra parte, en el proceso de evaluación.

Fig. 1. RECÁMARA PARA LA TOMA DE IMÁGENES



Fuente: Autores del proyecto

Fig. 2. ESTRUCTURA DE LA BASE DE DATOS PARA EL ALMACENAMIENTO DE IMÁGENES



Fuente: Autores del proyecto

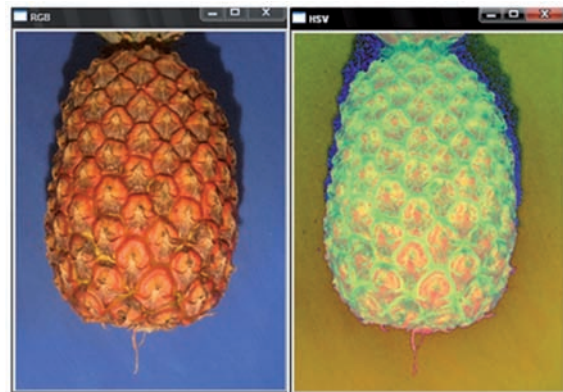
B. Pre-procesamiento de imágenes

En el pre-procesamiento de las imágenes, se realizaron las siguientes operaciones: cambio de las imágenes del espacio de color RGB al espacio de color HSV, obtención de los histogramas de los canales H y S, y filtrado de estos histogramas mediante un filtro de la mediana del tipo "moving average". La mayoría de estas operaciones se realizaron -con base en las librerías Open CV de Intel [26].

Las imágenes tomadas inicialmente se encuentran en el espacio de color RGB. Estas imágenes se transforman al espacio de color HSV con el fin de lograr una mejor separación y detección de los colores. En este espacio de color, cada color trabaja con 3 componentes básicas: matiz (Hue), saturación (Saturation) y luminancia (Value). El matiz hace referencia al valor de cromaticidad o clase de color. La saturación se refiere a las longitudes de onda que se suman a la frecuencia del color y determinan la cantidad de blanco que contiene un color. La luminancia se corresponde con la apreciación subjetiva de claridad y oscuridad [27].

El modelo de color HSV imita la manera como los inspectores de control de calidad de productos agrícolas determinan en forma manual el estado de madurez los mismos, mediante los cambios en el matiz y pureza del color. En la Fig. 3. se puede observar la transformación de una imagen de la piña del espacio RGB al espacio HSV.

Fig. 3. IMAGEN DE LA PIÑA EL ESPACIO RGB Y EN EL ESPACIO HSV



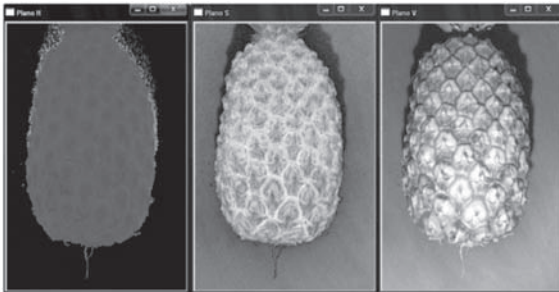
Fuente: Autores del proyecto

Una vez se tiene la imagen en el espacio de color HSV, se separa en los tres canales que la componen (Fig. 4). Se trabaja con los canales cromáticos, H y S, y se hallan sus respectivos his-

togramas. El histograma describe la probabilidad de distribución de los valores de intensidad en la imagen. Las distribuciones en los histogramas de los planos H y S están afectadas exteriormente por ruido, el cual se manifiesta en las variaciones bruscas del mismo. Con el fin de eliminar este ruido, se pueden aplicar filtros de suavizado, ya sea a cada píxel de la imagen, o a su correspondiente histograma. En este trabajo, se empleó un filtro de la mediana del tipo "moving average" [27] con una ventana de 9 aplicado tanto al histograma del plano H como al histograma de plano S. En la Fig. 5. se muestra el resultado de aplicar el filtro de suavizado al histograma del canal S de una imagen de la piña.

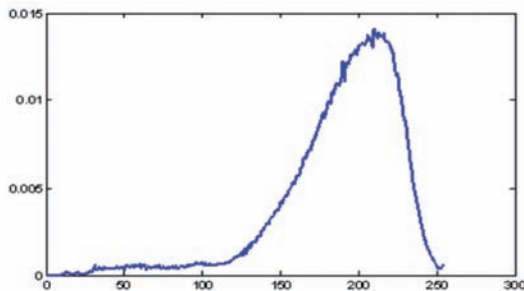
Se decidió aplicar el filtro a los dos histogramas porque las características empleadas en el proceso de clasificación son extraídas de los mismos y además, por la rapidez en el procesamiento pensando en una aplicación en línea. Una característica importante de este tipo de filtro es que al aplicarlo varias veces sobre el mismo histograma, se suavizan cada vez más las curvas de distribución y no se alteran significativamente los datos para su posterior utilización.

Fig. 4. IMAGEN DE LA PIÑA EN LOS PLANOS H, S Y V

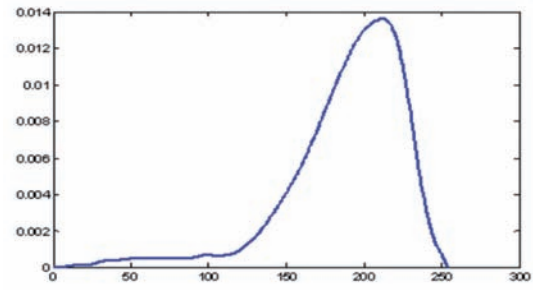


Fuente: Autores del proyecto

Fig. 5. FILTRADO DEL HISTOGRAMA DEL PLANO S: A) HISTOGRAMA SIN FILTRAR, B) HISTOGRAMA FILTRADO MEDIANTE EL FILTRO "MOVING AVERAGE"



A)



B)

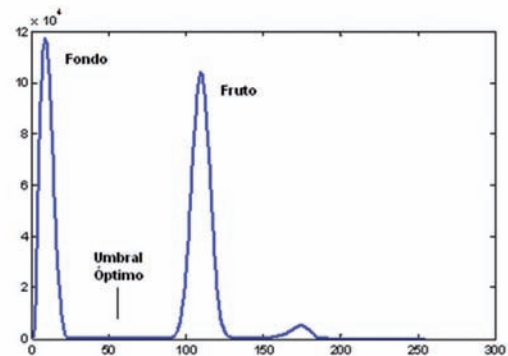
Fuente: Autores del proyecto

C. Segmentación

Del resultado de transformar la imagen al espacio de color HSV, se decide que el método de segmentación será por umbrales (thresholding) y usará la información de los histogramas del plano H o del plano S, ya que el plano V, equivale a una imagen resultante de haber transformado una imagen en el espacio RGB a una imagen en niveles de gris.

El análisis del comportamiento de los histogramas, determinó que la umbralización del histograma perteneciente al plano H era la mejor estrategia de segmentación. En la Fig. 6. se aprecia claramente la distribución normal que representa los píxeles pertenecientes al fondo y la distribución normal que representa los píxeles pertenecientes al fruto. La segmentación se consigue al barrer la imagen píxel a píxel y etiquetar cada píxel como "fruto" o "fondo", dependiendo de si el valor del píxel es mayor o inferior que un cierto valor (k) llamado umbral.

Fig. 6. HISTOGRAMA DEL PLANO H



Fuente: Autores del proyecto

La selección de los umbrales es un aspecto fundamental en el proceso de umbralización ya que determina en gran medida la calidad del resultado final de la segmentación. En este trabajo, la selección del umbral se realizó mediante el método de Otsu. Este método permite hallar el umbral óptimo entre dos distribuciones gaussianas de forma automática. Es un método muy empleado en aplicaciones industriales por su rapidez de ejecución y fiabilidad [24].

El método de Otsu se describe de la siguiente manera: Los píxeles de una imagen pueden ser representados por el conjunto de niveles de gris [1,2,..., L] donde L es el máximo nivel de intensidad. De la misma manera el número de píxeles en el nivel i puede ser denotado por n_i y la suma total de píxeles en la imagen puede ser denotada por N . Los niveles de gris en el histograma son normalizados y considerados como una distribución de probabilidad P_i .

$$P_i = \frac{n_i}{N} \quad (1)$$

Se asume que el fondo y el objeto (fruto) pertenecen a dos clases diferentes C_0 y C_1 , y que el umbral que las separa es k , las probabilidades de ocurrencia en cada clase serán:

$$\omega_0 = \sum_{i=1}^k P_i \quad y \quad \omega_1 = \sum_{i=k+1}^L P_i \quad (2)$$

Las medias y varianzas asociadas son:

$$\mu_0 = \sum_{i=1}^k \frac{i * P_i}{\omega_0} \quad y \quad \mu_1 = \sum_{i=k+1}^L \frac{i * P_i}{\omega_1} \quad (3)$$

$$\sigma_0^2 = \sum_{i=1}^k (i - \mu_0)^2 * \frac{P_i}{\omega_0} \quad y \quad (4)$$

$$\sigma_1^2 = \sum_{i=k+1}^L (i - \mu_1)^2 * \frac{P_i}{\omega_1}$$

Se obtiene finalmente la varianza ponderada,

$$\sigma_w^2 = \omega_0 * \sigma_0^2 + \omega_1 * \sigma_1^2 \quad (5)$$

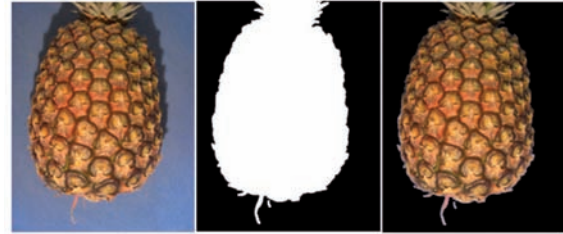
Según lo planteado por Otsu, en la medida en que las gaussianas se asemejan al histograma real, las desviaciones estándar tienden a disminuir y como consecuencia de ello, se debe elegir aquel umbral que minimice la suma de las varianzas de los dos lóbulos del histograma.

Una vez hallado el umbral óptimo, se procede a crear una imagen binaria a través de la cual, se extrae el área de interés de la imagen original. Todos los píxeles que están por encima del umbral óptimo se etiquetan como "1" y los píxeles que están por debajo de este umbral son etiquetados como "0". La imagen binaria y el área de interés como resultado del proceso de segmentación se muestran en la Fig. 7.

D. Extracción de características

Los momentos de color son una de las técnicas más utilizadas en el análisis de los histogramas para determinar el color predominante en una imagen.

Fig. 7. PROCESO DE SEGMENTACIÓN: IMAGEN ORIGINAL, IMAGEN BINARIA, IMAGEN SEGMENTADA



Fuente: Autores del proyecto

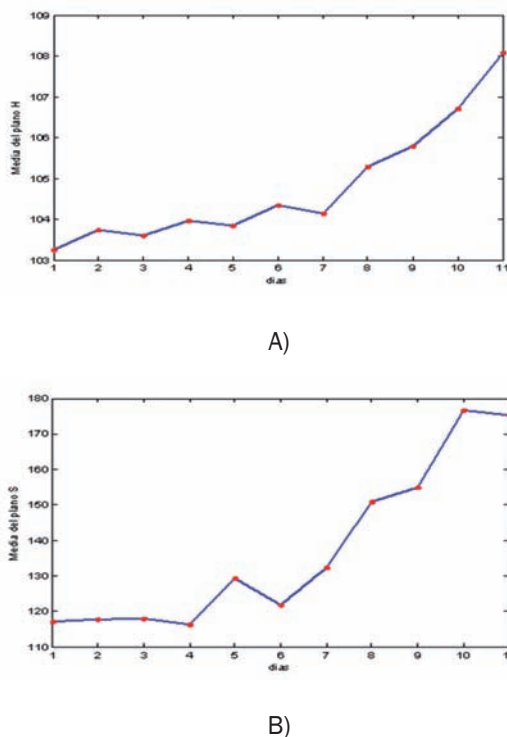
Según Stricker y Orengo [28], existen tres momentos centrales en la distribución del color en una imagen: la media que corresponde al primer momento, la desviación al segundo momento y la asimetría al tercer momento.

En este artículo, se seleccionó como características para extraer de cada imagen, el primer momento de distribución del color de los histogramas de los planos H y S. Para el histograma del plano H, la media indica el color dominante en la imagen y para el histograma del plano S, la media revela la cantidad de blanco que afecta la pureza del color. La media en cada histograma fue hallada con (3).

El cálculo de la media de los histogramas de los planos H y S se realizó para cada una de las imágenes tomadas a lo largo del proceso de maduración de cada piña. Estas imágenes estaban previamente almacenadas en la base de datos mostrada en la Fig. 2. A manera de ejemplo y para una piña en particular, en la Fig. 8. se observa la evolución de la media en cada componente de color (H y S) a medida que la piña madura. El eje

vertical representa la media del histograma para cada día y el eje horizontal el día en que fue tomada la imagen.

Fig. 8. EVOLUCIÓN DE LA MEDIA DEL HISTOGRAMAS DURANTE LOS DÍAS DEL PROCESO DE MADURACIÓN DE LA PIÑA: A) PLANO H, B) PLANO S



Fuente: Autores del proyecto

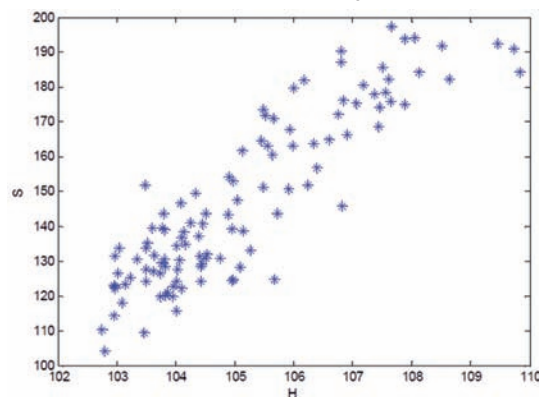
E. Clasificación

Este procedimiento permite clasificar las características extraídas y asignarlas dentro de una clase específica. Para este trabajo, se seleccionaron tres clases a priori denominadas piña verde, piña pintona y piña madura, las cuales corresponden a las tres etapas presentes en el proceso de maduración de la piña. Es preciso aclarar, que aunque se conoce el número de clases, son desconocidos los parámetros estadísticos que definen a cada una de ellas (media, varianza, correlación entre clases). Es por ello, que se utiliza el método de clasificación no supervisada por agrupamiento mediante clústers [25].

Las imágenes almacenadas en la base de datos de la Fig. 2. se dividen en dos conjuntos: de entrenamiento y de prueba. El clasificador se diseña con el conjunto de entrenamiento y luego se evalúa y se obtiene el error de clasificación para el conjunto de prueba.

El conjunto de entrenamiento está formado por 110 imágenes y las características extraídas son organizadas en dos vectores, uno para las medias de los histogramas de los planos H y el otro para las medias de los histogramas de los planos S. Estos vectores de características son graficados en el plano cartesiano como se observa en la Fig. 9. Es este plano, denominado espacio de características, el eje horizontal corresponde a los datos del vector de características del plano H y el eje vertical corresponde a los datos del vector de características del plano S.

Fig. 9. ESPACIO DE CARACTERÍSTICAS DEL CONJUNTO DE ENTRENAMIENTO



Fuente: Autores del proyecto

El algoritmo MBSAS es el encargado del aprendizaje en el clasificador. Consta de dos fases: la primera fase consiste en la determinación del centroide inicial de cada clúster, y en la segunda fase, los datos de los vectores de características sin asignar, son presentados por una segunda vez al algoritmo y son asignados al clúster apropiado, se empleó como criterio de agrupamiento la distancia mínima. Después de cada asignación, se recalcula el centroide correspondiente al clúster afectado.

El algoritmo MBSAS puede ser implementado mediante el siguiente pseudo código [25]:

Determinación de clústers

- $m=1$
- $C_m = \{x_i\}$
- For $i=2$ hasta N
 - Encontrar $C_k: d(x_i, c_k) = \min_{1 \leq j \leq m} d(x_i, c_j)$
 - If $(d(x_i, C_k) > \theta) \& (m < q)$ entonces
 - $m=m+1$
 - $C_m = \{x_i\}$
 - sino
 - $C_k = C_k \cup \{x_i\}$

- Donde sea necesario, actualizar los representantes
- Fin del {If}
- Fin del {For}

Clasificación de patrones

- For $i=2$ hasta N
 - If x_i aun no ha sido asignado a un cluster, entonces
 - Encontrar $C_k : d(x_i, C_k) = \min_{1 \leq j \leq m} d(x_i, C_j)$
 - $C_k = C_k \cup \{x_i\}$
 - Donde sea necesario, actualizar los representantes
- Fin del {If}
- Fin del {For}

Donde m es el número de clústers, C_m es el vector que representa los clústers, N es el tamaño del vector de datos, d es la distancia entre un clúster y un dato, C_k es el clúster al cual la distancia d es la más corta, θ es el umbral que representa una medida de proximidad entre las muestras, y q es el número máximo de clústers que el algoritmo puede formar.

Un factor importante que afecta los resultados del algoritmo de agrupamiento es la elección del umbral θ (parámetro de proximidad). El valor de este parámetro establece si una muestra pertenece o no a una clase determinada. Si el parámetro de proximidad hallado para una muestra dada es mayor que un valor previamente establecido, el algoritmo de agrupamiento es obligado a formar un nuevo clúster o clase. Además, este valor afecta la asignación de las muestras cercanas a los límites de cada clase, fenómeno que se conoce como traslape. En el presente trabajo, el parámetro de proximidad fue hallado por prueba y error de acuerdo al número de clases establecido y el plano de características de entrenamiento.

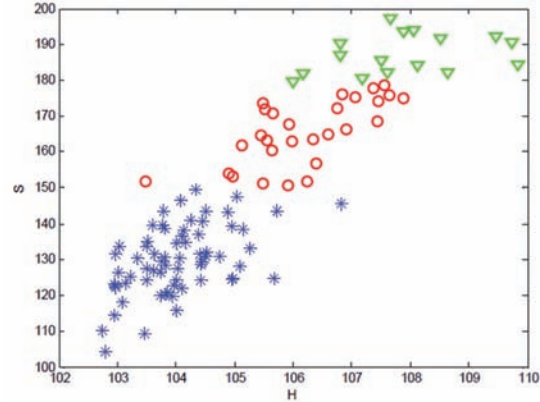
Al aplicar el algoritmo de agrupamiento previamente descrito, se obtuvo los resultados mostrados en la Fig. 10. Las muestras de color azul corresponden a la primera clase (piña verde), las muestras de color rojo corresponden a la segunda clase (piña pintona) y las muestras de color verde corresponden a la tercera clase (piña madura). Se observa que el criterio de agrupamiento afecta la asignación de las muestras cercanas a los límites de cada clase. Esto se debe a que los datos en el eje H están muy cercanos entre sí (intervalo entre 102 - 110) comparados con los datos en el S (in-

tervalo entre 100 - 200). Para dar una mejor distribución de las clases, fue necesario escalar los datos del eje H mediante la siguiente expresión:

$$H_{new} = 100 \ln(H-100) \quad (6)$$

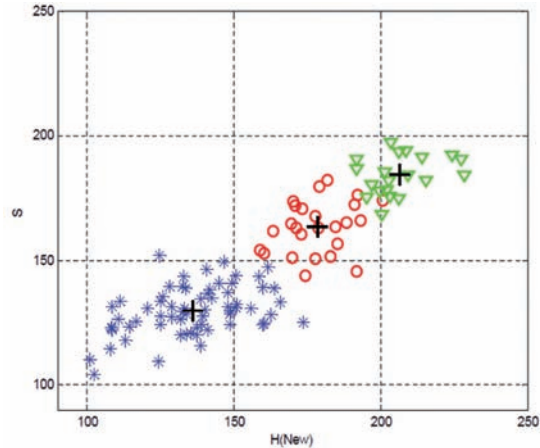
Los tres centroides obtenidos en la fase de entrenamiento se muestran en la Fig. 11. Estos centroides definen las tres etapas de maduración de la piña y serán utilizados posteriormente por el clasificador en la etapa de evaluación.

Fig. 10. AGRUPAMIENTO DE LAS CARACTERÍSTICAS DEL CONJUNTO DE ENTRENAMIENTO



Fuente: Autores del proyecto

Fig. 11. CENTROIDES FINALES DE LA ETAPA DE ENTRENAMIENTO



Fuente: Autores del proyecto

F. Evaluación

La evaluación del desempeño se basa en la estimación del error de clasificación. Con base en el error obtenido se puede predecir el desempeño del clasificador ante nuevas muestras.

Para evaluar el desempeño del clasificador se realizó validación cruzada [29] para un conjunto de prueba de 400 imágenes, que corresponden

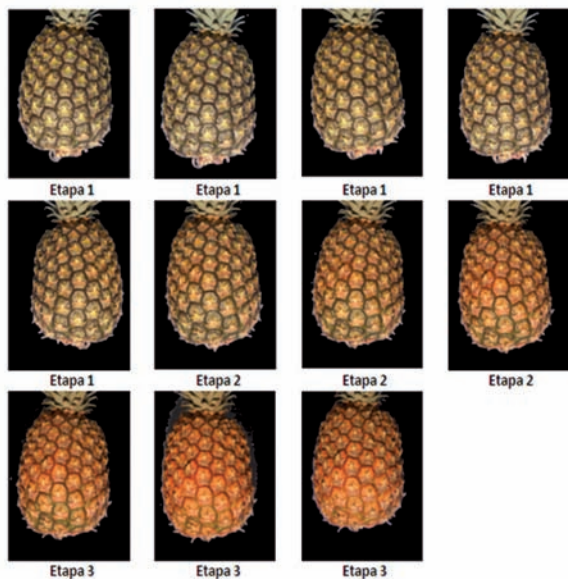
a imágenes de diferentes piñas en diferentes etapas de maduración.

El clasificador encuentra la distancia euclídea mínima entre las características de cada muestra de evaluación y los centroides obtenidos en la etapa de entrenamiento. Una vez hallada la distancia mínima, se determina a cual clúster pertenece la muestra de evaluación lo que define el estado de maduración en que se encuentra la piña.

A manera de ejemplo, en la Fig. 12. se muestra los resultados de clasificación obtenidos para un conjunto de imágenes que describen la evolución en el proceso de maduración de una piña perteneciente al conjunto de evaluación. Se ha llamado Etapa 1 a la muestra que pertenece a la clase piña verde, Etapa 2 a la muestra que pertenece a la clase piña pintona y Etapa 3 a la muestra que pertenece a la clase piña madura. Resultados adicionales del proceso de evaluación pueden ser encontrados en [30].

Con el fin de evaluar la fiabilidad del clasificador, se decidió contrastar los resultados dados por el clasificador propuesto en este artículo con el criterio de clasificación manual dado por tres expertos. Los expertos son personas con más de 10 años de experiencia en el cultivo y comercialización de la piña.

Fig. 12. RESULTADO DE CLASIFICACIÓN EN EL PROCESO DE EVALUACIÓN PARA UNA PIÑA DE EJEMPLO



Fuente: Autores del proyecto

Para ello, del conjunto de evaluación se seleccionaron las imágenes tomadas a lo largo del pro-

ceso de maduración de 5 piñas. Esto corresponde a la evaluación de 55 imágenes, tanto por parte del clasificador, como por parte de cada uno de los expertos. Ellos, basados en su experiencia, determinaron a cuál etapa de maduración pertenecía cada una de las muestras

Los resultados de clasificación manual de los tres expertos y los resultados del clasificador propuesto en este artículo, se presentan en las Tablas 1, 2, 3, 4 y 5. El resultado, 1, 2 o 3 corresponde a cada una de las etapas de clasificación relacionadas anteriormente.

Como se observa en la Tabla I, tanto los expertos como el clasificador determinan que la piña 1 está clasificada correctamente. Sin embargo en la Tabla II, los expertos afirman que en el día 8, la piña 2 aún está en la etapa 2 mientras que el clasificador la ubica en la etapa 3. En la Tabla III, el experto 3 asegura que para el día 6, la piña 3 está en la etapa 1, mientras que el clasificador y los otros expertos la ubican en la etapa 2. En la Tabla IV, el mismo experto afirma que en el día 6 la piña 4 aún se encuentra en la etapa 2, sin embargo los otros dos expertos y el clasificador la ubican en la etapa 1. Para la piña 5, los expertos 1 y 2 determinan que en el día 8 aún está en etapa 2, sin embargo en la Tabla V, se observa que el experto 3 está de acuerdo con el clasificador.

TABLA I

RESULTADOS DE CLASIFICACIÓN AUTOMÁTICA Y CLASIFICACIÓN MANUAL PARA LA PIÑA 1

Días	1	2	3	4	5	6	7	8	9	10	11
Clasificador	1	1	1	1	1	2	2	2	3	3	3
Experto Nro. 1	1	1	1	1	1	2	2	2	3	3	3
Experto Nro. 2	1	1	1	1	1	2	2	2	3	3	3
Experto Nro. 3	1	1	1	1	1	2	2	2	3	3	3

Fuente: Autores del proyecto

TABLA II

RESULTADOS DE CLASIFICACIÓN AUTOMÁTICA Y CLASIFICACIÓN MANUAL PARA LA PIÑA 2

Días	1	2	3	4	5	6	7	8	9	10	11
Clasificador	1	1	1	1	1	2	2	3	3	3	3
Experto Nro. 1	1	1	1	1	1	2	2	2	3	3	3
Experto Nro. 2	1	1	1	1	1	2	2	2	3	3	3
Experto Nro. 3	1	1	1	1	1	2	2	2	3	3	3

Fuente: Autores del proyecto

TABLA III
RESULTADOS DE CLASIFICACIÓN AUTOMÁTICA Y CLASIFICACIÓN MANUAL PARA LA PIÑA 3

Días	1	2	3	4	5	6	7	8	9	10	11
Clasificador	1	1	1	1	1	2	2	2	3	3	3
Experto Nro. 1	1	1	1	1	1	2	2	2	3	3	3
Experto Nro. 2	1	1	1	1	1	2	2	2	3	3	3
Experto Nro. 3	1	1	1	1	1	1	2	2	3	3	3

Fuente: Autores del proyecto

TABLA IV
RESULTADOS DE CLASIFICACIÓN AUTOMÁTICA Y CLASIFICACIÓN MANUAL PARA LA PIÑA 4

Días	1	2	3	4	5	6	7	8	9	10	11
Clasificador	1	1	1	1	1	1	2	2	3	3	3
Experto Nro. 1	1	1	1	1	1	1	2	2	3	3	3
Experto Nro. 2	1	1	1	1	1	1	2	2	3	3	3
Experto Nro. 3	1	1	1	1	1	2	2	2	3	3	3

Fuente: Autores del proyecto

TABLA V
RESULTADOS DE CLASIFICACIÓN AUTOMÁTICA Y CLASIFICACIÓN MANUAL PARA LA PIÑA 5

Días	1	2	3	4	5	6	7	8	9	10	11
Clasificador	1	1	1	1	1	1	2	3	3	3	3
Experto Nro. 1	1	1	1	1	1	1	2	2	3	3	3
Experto Nro. 2	1	1	1	1	1	1	2	2	3	3	3
Experto Nro. 3	1	1	1	1	1	1	2	3	3	3	3

Fuente: Autores del proyecto

De acuerdo a estos resultados, se concluye que los expertos tienen criterios diferentes en determinadas muestras. En la mayoría de los casos, los expertos tienen sus diferencias en la clasificación en los límites de transición de cada etapa de maduración.

Con el fin de cuantificar el error de clasificación y medir la fiabilidad del clasificador propuesto en este artículo con respecto al criterio de clasificación manual de los expertos, se realizó una estimación del error de clasificación para cada etapa de maduración de la piña, mediante la siguiente expresión:

Donde,

$$\%Error = \left| \frac{MTC - MCC}{MTC} \right| * 100\% \quad (7)$$

MTC es el número de muestras totales por etapa clasificadas por el algoritmo propuesto.

MCC es el número de muestras por etapa clasificadas por el algoritmo propuesto y que coinciden con el criterio de los expertos.

Con los resultados de clasificación automática y de clasificación manual mostrados en las Tablas 1, 2, 3, 4 y 5, los porcentajes de error de clasificación para cada etapa de maduración se presentan en la Tabla VI. Se asume como muestra no clasificada correctamente, a aquella muestra cuya clasificación dada por el algoritmo de clasificación automática, no coincide con el criterio de clasificación manual dado por los expertos.

TABLA VI
ERROR DE CLASIFICACIÓN PARA CADA ETAPA DE MADURACIÓN DE LA PIÑA

Etapa	Muestras totales	Muestras no clasificadas correctamente	% Error
1	27	0	0
2	11	0	0
3	17	2	11,76
Total	55	2	3,63

Fuente: Autores del proyecto

El mayor porcentaje de error se da en la etapa 3. Para determinar la causa del error, se analizan los agrupamientos finales del conjunto de entrenamiento mostrados en la Fig. 11. Se aprecia claramente un pequeño traslape entre los límites de las etapas 2 y 3, lo que introduce cierto grado de incertidumbre en el proceso de clasificación en muestras ubicadas en estos límites.

En general, el clasificador automático presenta un error de clasificación de 3,64%, lo que indica que el proceso de clasificación tiene un desempeño del 96,36%. Este porcentaje de clasificación garantiza que los resultados obtenidos son eficientes y confiables.

III. CONCLUSIONES

En el artículo, se ha presentado el desarrollo de un algoritmo que permite identificar las etapas de maduración de la piña en su variedad perolera. El algoritmo implementa cada una de las etapas de un sistema de visión artificial: adquisición de imágenes, pre-procesamiento, segmentación, extracción de características, clasificación y evaluación.

La segmentación se realiza a partir del método de umbralización del histograma de la componente H del espacio de color HSV. El umbral es hallado automáticamente a través del método de Otsu. El

vector de características está formado por el primer momento del color de las distribuciones del histograma de las componentes H y S. El clasificador emplea el algoritmo MBSAS y permite identificar tres etapas de grado de maduración: piña verde, piña pintona y piña madura.

La efectividad del algoritmo de clasificación fue del 96,36%, el cual fue evaluado según el criterio de clasificación manual de tres expertos. La aplicación del clasificador presentado en este trabajo puede ser extendida a diferentes productos agrícolas de la región. Como trabajos futuros, se plantea el desarrollo de nuevas estrategias para el entrenamiento del clasificador con el fin de disminuir el error de clasificación.

REFERENCIAS

- [1] Colombia ha tenido un buen año con los TLC que negoció, Revista Portafolio (Online), Octubre 13 de 2011, Disponible: <http://www.portafolio.co/negocio>
- [2] El país está preparado agrícolamente para el TLC, Revista Portafolio (Online), Octubre 11 de 2011, Disponible: <http://www.portafolio.co/negocio>
- [3] R. D. Tillett, Image analysis for agricultural processes. En: Division Note DN 1585. Research Institute. 1990.
- [4] L.G. Shapiro, G.C. Stockman, Computer Vision, Prentice Hall, 2001.
- [5] Da-Wen Sun, Computer Vision Technology for Food Quality Evaluation, Ed. Elsevier, 2008.
- [6] P. Chen, Z. Sun, A review of non-destructive methods for quality evaluation and sorting of agricultural products, Journal of Agricultural Engineering Research, Volume 49, May–August 1991, Pages 85-98.
- [7] Tadhg Brosnan, Da-Wen Sun, Inspection and grading of agricultural and food products by computer vision systems—a review, Computers and Electronics in Agriculture, Volume 36, Issues 2–3, November 2002, Pages 193-213.
- [8] V Leemans, M-F Destain, A real-time grading method of apples based on features extracted from defects, Journal of Food Engineering, 61 (2004), pp. 83–89.
- [9] JA Throop, DJ Aneshansley, WC Anger, DL Peterson, Quality evaluation of apples based on surface defects: development of an automated inspection system, Postharvest Biology and Technology, 36 (2005), pp. 281–290.
- [10] D Unay, B Gosselin, Stem and calyx recognition on “Jonagold” apples by pattern recognition, Journal of Food Engineering, 78 (2) (2007), pp. 597–605.
- [11] J Blasco, S Cubero, R Arias, F Juste, E Moltó, On-line quality grading of mandarin segments by computer vision, XVIth CIGR World Congress: Agricultural Engineering for a Better World, Book of Abstracts, Bonn, Germany (2006), pp. 623–624.
- [12] J Blasco, N Aleixos, E Moltó, Computer vision detection of peel defects in citrus by means of a region oriented segmentation algorithm, Journal of Food Engineering, 81 (3) (2007), pp. 535–543.
- [13] N Hernandez, P Barreiro, M Ruiz-Altisent, J Ruiz-Cabello, ME Fernandez-Valle, Detection of seeds in citrus using MRI under motion conditions and improvement with motion correction, Concepts in Magnetic Resonance: Part B—Magnetic Resonance Engineering, 26B (1) (2005), pp. 81–92.
- [14] M Nagata, PM Bato, M Mitarai, Q Cao, T Kitahara, Study on sorting system for strawberry using machine vision (part 1): development of software for determining the direction of strawberry (Akihime variety), Journal of Japan Society of Agricultural Machinery, 62 (1) (2000), pp. 100–110.
- [15] R Díaz, M Blasco, J Blasco, E Moltó, Comparison of three algorithms in the classification of table olives by means of computer vision, Journal of Food Engineering, 61 (2004), pp. 101–107.
- [16] S Colmagro, G Collins, M Sedgley, Processing technology of the table olive, Horticultural Reviews, 25 (2001), pp. 235–260.
- [17] JC Noordam, GW Otten, AJM Timmermans, BH van Zwol, High speed potato grading and quality inspection based on a color vision system, Proceedings of SPIE: The International Society for Optical Engineering, 3966 (2000), pp. 206–219.
- [18] T Brosnan, D Sun, Improving quality inspection of food products by computer vision—a review, Journal of Food Engineering, 61 (2004), pp. 125–135.

- [19] C. Maya. Desarrollo de un algoritmo para la caracterización y clasificación de granos de café empleando técnicas de visión artificial. Manizales: U. Nacional de Manizales, 2001.
- [20] J. Betancur, Segmentación de granos de café mediante métodos de crecimiento de regiones. Manizales: U Nacional de Manizales, 2002.
- [21] Z. Sandoval, F. Prieto, Caracterización y clasificación de café cereza por medio de visión artificial, 2003.
- [22] J. E. Hernández, Clasificación de frutos de café según su etapa de maduración basada en redes neuronales artificiales. Manizales: U. Nacional de Manizales, 2004.
- [23] J. C. Mosquera, A. Sepúlveda, C, A. Isaza, Procesamiento de imágenes ópticas de frutos café en cereza por medio de filtros acusto-ópticos, Ingeniería y Desarrollo, Universidad del Norte, Numero 21, 2007.
- [24] N.Otsu, A threshold selection method for gray level histograms, IEEE Trans. Syst. Man Cybern., 9, 62-6, 1979.
- [25] S. Theodoridis, S. Koutroumbas, Pattern Recognition. Academic Press, San Diego, USA, 2008.
- [26] G. Bradski, A. Kaehler, Learning OpenCV: Computer Vision with the OpenCV Library, O'Reilly Media, 2008.
- [27] González, R.C., Woods, R.E.: Digital Image Processing. Prentice Hall, 2008.
- [28] Stricker, M. and Orengo, M., Similarity of color images, SPIE Conference on Storage and Retrieval for Image and Video Databases III. Vol. 2420, p. 381-392, 1995.
- [29] Moore, A., Cross validation for detecting and preventing overfitting, School of Computer Science, Carnegie Mellon University, 2001.
- [30] Lizcano, S., Identificación de las etapas de maduración de la piña perolera empleando técnicas de visión artificial, Universidad Pontificia Bolivariana, Seccional Bucaramanga, 2010.