

Nueva técnica para comprobar la calidad del arroz fundamentada en el tratamiento de imágenes

Julio Cesar Romero Pabon

Universidad del Atlántico, Barranquilla (Colombia)

julioromero@mail.uniatlantico.edu.co

Resumen

En este trabajo se presenta una técnica basada en el análisis matemático y el tratamiento de imágenes para determinar el porcentaje de granos partidos en una muestra de arroz. El objetivo de esta técnica es que, a pesar de las condiciones tan diferentes entre las muestras (tipo de iluminación, características del grano, ausencia o presencia de ruido), esta es capaz de extraer correctamente “estructuras relevantes” para posteriormente, realizar sobre ellas la extracción y el análisis de sus características. Para lograr esto, se diseñó e implementó un algoritmo basado en esta nueva técnica para determinar la calidad del arroz.

Palabras clave: segmentación, análisis, matemática, umbralización, cuantización.

New technique to check the quality of rice based on image treatment

Abstract

This work presents a technique based on the mathematical analysis and the treatment of images to determine the percentage of broken grains in a rice sample. The objective of this technique is that, in spite of the different conditions between the samples (type of illumination, grain characteristics, absence or presence of noise), it is able to correctly extract "relevant structures" Extraction and analysis of its characteristics. To achieve this, an algorithm based on this new technique was designed and implemented to determine rice quality.

Key words: segmentation, analysis, mathematics, thresholding, quantization.

1. INTRODUCCIÓN

Este trabajo describe un sistema de procesamiento de imágenes digitales para determinar el número de granos partidos y enteros en muestras de arroz, factor de mucha importancia en la calidad del mismo para su comercialización. El principal objetivo es contar el número de granos (partidos y no partidos) en una imagen, proveniente de un escáner o cámara fotográfica, disminuir el margen de error en estos procesos para poder hacer una buena inferencias sobre el porcentaje de granos partidos. Con este fin se ha diseñado una técnica basada en las matemáticas y el tratamiento de imágenes con un algoritmo eficiente para realizar una excelente segmentación de la imagen que contiene las muestras de arroz.

La segmentación de una imagen consiste en extraer propiedades o características comunes de alguna región de interés. La segmentación por histograma realizada en este trabajo consiste en retener aquellos píxeles que se encuentren en un cierto entorno de

niveles de color del histograma definido por el usuario. Es importante resaltar que todo sistema de visión que use segmentación necesita realizar los siguientes pasos: tener una imagen digital, realizar un filtrado a la imagen, segmentar la imagen, extraer y analizar las características capturadas.

Uno de los primeros productos de consumo a nivel mundial es el arroz y su comercialización está regida por unos estándares internacionales de calidad, en los cuales se contempla que el valor del grano está en función de la cantidad de granos partidos. Esta labor de clasificación de los granos de arroz hasta el momento es realizada de forma tradicional (rudimentaria o artesanal) en la mayoría de las empresas que lo comercializan. El proceso para la clasificación y control de calidad del grano de arroz que se lleva a cabo en las empresas comercializadoras de arroz es realizado en forma manual, para lo cual, se toman varias muestras de arroz para su clasificación; esta son observadas y valoradas por una persona que generalmente no utiliza ninguna herramienta mecánica ni computacional. El objetivo de este trabajo es diseñar e implementar una nueva técnica basas en las matemáticas y en el tratamiento de imágenes para determinar el porcentaje de granos partidos en muestras de arroz, e manera que se agilicen y fortalezcan los procesos de clasificación en cuanto el número de granos partidos, con más precisión y auditoria para estos procesos.

2. FUNDAMENTOS TEÓRICOS DE MATEMÁTICA Y DEL TRATAMIENTO DE IMÁGENES

Los conceptos básicos empleados para desarrollar la herramienta computacional propuesta como una solución a la problemática planteada. Para diseño y elaboración del algoritmo para el tratamiento de imagen es necesario tener claro algunos temas como son: la imagen digital, análisis del color, los métodos de segmentación basados en umbralización, crecimiento de regiones, extracción de características en un mapa de bit y la aplicación estadística.

2.1 IMAGEN DIGITAL O DE BITS

En informática, conjunto secuencial de bits que representa, en la memoria, una imagen que puede visualizarse en la pantalla, especialmente en los sistemas que disponen de interfaz gráfica de usuario. Cada bit de una imagen de bits corresponde a un píxel en la pantalla. Por ejemplo, la propia pantalla representa una única imagen de bits; de forma similar, las configuraciones de puntos para todos los caracteres de una fuente representan la imagen de bits de dicha fuente. En equipos informáticos con pantalla en blanco y negro, los valores de bit de una imagen de bits pueden ser 0 para blanco, o 1 para negro. La configuración de ceros y unos en la imagen de bits determina la configuración de los puntos blancos y negros que forman una imagen en la pantalla. En los equipos dotados de color, la descripción correspondiente de los bits en pantalla se denomina imagen de píxel, ya que es necesario más de un bit para representar cada píxel. En los actuales PC se pueden representar combinaciones de 24 bits de color (millones de colores).

2.2 PERMUTACIONES O NÚMERO DE COLORES EN UN MAPA DE BITS

La permutaciones se obtienen al analizar el color que contiene cada pixel, estas permutaciones dentro de los RGB arrojan un código o número que identifican el color de cada pixel, de las diferentes composiciones obtenidas por medio de los RGB para la determinación específica de un color en la imagen, es posible formar los diferentes conjuntos de pixeles que permitirán aplicar los conceptos matemáticos como son frontera de una región, punto interior, homogeneidad y conectividad entre conjuntos. La descripción **RGB** (del inglés *Red, Green, Blue*; "rojo, verde, azul") de un [color](#) hace referencia a la composición del color en términos de la intensidad de los [colores primarios](#) con que se forma: el rojo, el verde y el azul. Es un modelo de color basado en la [síntesis aditiva](#), con el que es posible representar un color mediante la mezcla por adición de los tres colores luz primarios.

Para indicar con qué proporción mezclamos cada color, se asigna un valor a cada uno de los colores primarios, de manera, por ejemplo, el valor 0 significa que no interviene en la mezcla, y a medida que ese valor aumenta, se entiende que aporta más intensidad a la mezcla. Aunque el intervalo de valores podría ser cualquiera (valores reales entre 0 y 1, valores enteros entre 0 y 37, etc.), es frecuente que cada color primario se codifique con un [byte](#) (8 [bits](#)). Así, de una manera estándar, la intensidad de cada una de las componentes se mide según una escala que va del 0 al 255. El número de colores diferentes que se pueden obtener en el computador, basado en los RGB, estará dada por la permutación: $P(\text{colores diferentes}) = R * G * B = 256 * 256 * 256 = 16'777.216$

2.3 HISTOGRAMA

El histograma de una imagen es una función discreta que representa el número de píxeles en la imagen en función de los niveles de intensidad. La probabilidad de ocurrencia de un determinado nivel se define como: $P(g) = \frac{N(g)}{M}$. Donde M es el número de píxeles en la imagen, $N(g)$ es el número de píxeles con el nivel de intensidad g .

2.4 SEGMENTACIÓN

Consiste en dividir una imagen digital en regiones o entidades significativas, esto es, tomar las partes o segmentos que se pueden considerar como unidades homogéneas relevantes, con respecto a una o más características. El objetivo es etiquetar las diferentes regiones que identifican a los objetos de interés. Todo esto se puede expresar matemáticamente, es decir, si R es la región que incluye la imagen completa, podemos definir la segmentación como un proceso que divide a R en k subregiones o subconjuntos disjuntos no vacíos $R_1, R_2, R_3, \dots, R_k$ cumpliéndose que

- a) $\cup R_i = R$ para $i = 1, 2, 3, \dots, k$
- b) R_i es una región conectada; para $i = 1, 2, 3, \dots, k$

- c) $R_i \cap R_j = \emptyset$ para todo i y j , $i \neq j$
- d) $P(R_i) = \text{verdadero}$ para $i = 1, 2, 3, \dots, n$ y
- e) $P(R_i \cap R_j) = \text{falso}$ para $i \neq j$

Donde $P(R_i)$ es un predicado lógico sobre los puntos del conjunto R_i . Un algoritmo de segmentación debe cumplir con estas condiciones.

La unidad que rige a la segmentación suele corresponder a píxeles, regiones o contornos que muestran una similitud por alguna característica en particular. Las características comúnmente utilizadas para segmentar regiones son: intensidad, textura, color, gradiente, movimiento, profundidad relativa entre otras.

2.5. EXTRACCIÓN DE CARACTERÍSTICAS

Una vez segmentada una imagen es importante considerar la forma de describir los objetos que aparecen después de realizar el proceso de segmentación. Las características son usadas para describir un objeto o sus atributos. La extracción de las características es el paso previo a la clasificación. Las características pueden ser extraídas por las propiedades topológicas, métricas, de irregularidad, por el esqueleto de la región y por características de textura de la región.

2.5.1. PROPIEDADES TOPOLÓGICAS

La característica principal de las propiedades topológicas es que no dependen de las variaciones que pueden presentar las regiones y no involucran nociones de distancia. Una de las propiedades topológicas de las regiones, más utilizadas es el número de componentes conectados. Un componente conectado es un subconjunto de dimensión máxima tal que dos puntos cualesquiera pueden unirse por una curva continua perteneciente al subconjunto. Otra de las propiedades usadas es el número de huecos en la región.

El número de huecos en la figura es igual al número de componentes conectados en el complemento de la figura.

Punto interior: Sea $S \subseteq M$, un punto a de S se llama *punto interior* de S si alguna de las bolas $B_M(a; r)$ está contenida en S . El interior de S , $\text{int } S$, es el conjunto de los puntos interiores a S . Un conjunto S es *abierto* en M si todos sus puntos son interiores; es *cerrado* en M si $M - S$ es abierto en M .

Frontera de un conjunto: Sea S un subconjunto de un espacio métrico M . Un punto x de M se llama punto de frontera de S si cada bola $B_M(x; r)$ contiene por lo menos, un punto de S y, por lo menos, un punto de $M - S$. El conjunto de todos los puntos de frontera de S se llama frontera de S y se designa por ∂S .

2.5.2. PROPIEDADES MÉTRICAS

Las propiedades métricas son generalizaciones de la distancia euclidiana. Entre las propiedades métricas más comunes encontramos el área, el perímetro y el centro de gravedad. El área A es el número de píxeles contenidos dentro de los límites de la región y esta dado por $A = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} g(i, j)$, el perímetro que es la medida en píxeles de la línea de contorno o los límites de una región.

Los códigos de cadena se utilizan para representar un conjunto de puntos, que constituyen una línea recta o no, también se usan para representar una frontera como un conjunto de segmentos con longitud y dirección. Para la definición de códigos de cadena, se tiene en cuenta la localización de un píxel (i, j) y sus 8 vecinos en las direcciones cuantizadas de 45° . A cada dirección se le asigna un valor numérico, así para las direcciones O, SO, S, SE, E, NE, N y NO les corresponden los valores 5, 6, 7, 8, 1, 2, 3 y 4

respectivamente, es lo que se conoce como código de cadena 8-direccional. (Fig. 1).

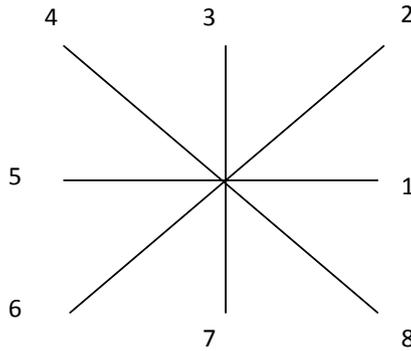


Figura 1. Código de cadena 8

El perímetro, otra de las propiedades métricas, es la longitud del borde de la región y se obtiene a partir del código de cadena de la frontera por: $P = \sum_i \sqrt{(X_i - X_{i-1})^2 + (Y_i - Y_{i-1})^2}$.

El centro de gravedad de la región es un único punto representativo de la región, tomando en cuenta el valor de la intensidad en cada punto, éste se obtiene de la siguiente manera: $X = \frac{\sum_i f(x_i, y_i) x_i}{\sum_i f(x_i, y_i)}$, $Y = \frac{\sum_i f(x_i, y_i) y_i}{\sum_i f(x_i, y_i)}$. Otra propiedad métrica son los ejes mayor y menor de una región, empleados para obtener la orientación de la región. La orientación de la región se obtiene del conjunto de ecuaciones:

$$S_x = \sum x_i, \quad S_y = \sum y_i, \quad S_{xx} = \sum x_i^2, \quad S_{yy} = \sum y_i^2, \quad S_{xy} = \sum x_i y_i \quad (5)$$

$$M_{xx} = S_{xx} - \frac{S_x^2}{A}, \quad M_{yy} = S_{yy} - \frac{S_y^2}{A}, \quad M_{xy} = S_{xy} - \frac{S_x S_y}{A}, \quad (6)$$

$$\emptyset = \tan^{-1} \left(\frac{M_{xx} - M_{yy} + \sqrt{(M_{xx} - M_{yy})^2 + 4 M_{xy}}}{2 M_{xy}} \right) \quad (7)$$

2.6 CALIDAD DEL ARROZ

Existen más de 2000 variedades de arroz cultivadas en el mundo. El banco de genes del IRRI en las Filipinas conserva no obstante alrededor de 83 000 variedades. Las diferencias se refieren a la morfología de la planta y del grano, la calidad del grano, la resistencia al acampamiento, la precocidad, la ramificación, la resistencia y tolerancia a los factores bióticos (malezas, insectos y enfermedades) y abióticos (frío, sequía, acidez del suelo, carencias en elementos minerales primordiales, etc.) y la productividad física.

- ✓ El mercado de arroz de calidad superior con una baja tasa de quebrados (menos de 10%) está dominado por la Tailandia, el Vietnam y los Estados Unidos, y responde principalmente a la demanda de los países industrializados.
- ✓ El mercado de arroz de calidad inferior (más de 10% de quebrados) está dominado por los países asiáticos (Tailandia, Vietnam e India) y responde principalmente a la demanda de los países en desarrollo de África, América latina, o de Asia.
- ✓ Calidad media: 15/20 % de quebrados.
- ✓ Calidad baja: 25/35 % de quebrados y hasta 100% de quebrados.

Toda irregularidad (decoloración o afrecho residual) baja el grado de calidad

2.7 ESTADÍSTICOS

Para estimar el porcentaje de granos partidos de la población de arroz, se hará uso del estimador de proporción, |debido a que los

experimentos realizados sobre las muestras de arroz cumplen con las condiciones binomiales (éxito y fracaso en n pruebas).

ESTIMACIÓN DE UNA PROPORCIÓN

Un estimador puntual de la proporción p en un experimento binomial está dado por el estadístico $\hat{p} = \frac{X}{n}$ donde X representa el número de éxitos en n pruebas. Por lo tanto, la proporción de la muestra $\hat{p} = \frac{x}{n}$ se utilizará como el estimador puntual del parámetro p . Si no se espera que la proporción p desconocida esté demasiado cerca de cero o de 1, podemos establecer un intervalo de confianza para p al considerar la distribución muestral de \hat{p} . Al designar un fracaso en cada prueba binomial con el valor 0 y un éxito con el valor 1, el número de éxitos, x , se puede interpretar como la suma de n valores que consisten sólo de ceros y unos, y \hat{p} es sólo la media muestral de estos n valores. De aquí, por el teorema del límite central, para n suficientemente grande, \hat{p} está distribuida de forma aproximadamente normal con media

$$\mu_{\hat{p}} = E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{np}{n} = p. \text{ Y } \text{varianza}$$

$$\sigma_{\hat{p}}^2 = \sigma_{\frac{X}{n}}^2 = \frac{\sigma_X^2}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}. \text{ Por tanto, podemos asegurar que}$$

$$P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha. \text{ Donde, } Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}, \text{ y } z_{\frac{\alpha}{2}} \text{ es el}$$

valor de la curva normal estándar sobre la cual encontramos un área de $\frac{\alpha}{2}$. Al sustituir para Z , escribimos

$$P\left(-z_{\frac{\alpha}{2}} < \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} < z_{\frac{\alpha}{2}}\right) = 1 - \alpha. \text{ Al multiplicar cada término de}$$

la desigualdad por $\sqrt{\frac{pq}{n}}$, y después restar \hat{p} y multiplicar por -1 ,

$$\text{obtenemos: } P\left(\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}\right) = 1 - \alpha. \text{ Es}$$

difícil manipular las desigualdades de manera que se obtenga un intervalo aleatorio cuyos puntos extremos sean independientes de p , el parámetro desconocido. Cuando n es grande, se introducen errores muy pequeños al sustituir la estimación puntual $\hat{p} = \frac{x}{n}$ por p bajo el signo del radical. Entonces podemos

escribir $P\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \approx 1 - \alpha$. Para

nuestra muestra aleatoria particular de tamaño n , se calcula la proporción muestral $\hat{p} = \frac{x}{n}$ y se obtiene el siguiente intervalo de confianza de $(1-\alpha)100\%$ aproximado para p . Intervalo de confianza de p de una muestra grande. Si \hat{p} es la proporción de éxitos en una muestra aleatoria de tamaño n , y $\hat{q} = 1 - \hat{p}$, un intervalo de confianza aproximado de $(1-\alpha)100\%$ para el parámetro binomial p está dado por

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

Donde $z_{\frac{\alpha}{2}}$ es el valor z que deja un área de $\frac{\alpha}{2}$ a la derecha.

Cuando n es pequeña y la proporción desconocida p se considera cercana a 0 o a 1, el procedimiento del intervalo de confianza que se establece aquí no es confiable y, por lo tanto, no se debería emplear. Para estar seguro, se requiere que tanto $n\hat{p}$ como $n\hat{q}$ sean mayores que o iguales a 5. El método para encontrar un intervalo de confianza para el parámetro binomial p también se aplica cuando la distribución binomial se utiliza para aproximar la distribución hipergeométrica; es decir, cuando n es pequeña en relación con N . El error de estimación será la diferencia absoluta entre \hat{p} y p , y podemos tener el nivel

de confianza de que esta diferencia no excederá $z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$.

3. METODOLOGÍA PARA EL DISEÑO Y ELABORACIÓN DE LA TÉCNICA QUE DETERMINA LA CALIDAD DEL

ARROZ BASADA EN LAS MATEMÁTICAS Y EL TRATAMIENTO DE IMÁGENES

Para analizar el componente matemático de una imagen digital, es necesario llevarla a matrices numéricas basadas en los espectros físicos del color, los RGB (sigla en inglés de red, green, blue, en español «rojo, verde y azul»), que representan la composición del color en términos de la intensidad de los colores primarios de la luz. La detección de las regiones de interés o donde están presente los granos de arroz en la imagen, se realiza por la aplicación de un filtro basado en el análisis matemático y ecuaciones diferenciales, ya que para la detección de bordes se hace uso del concepto matemático de frontera, y para diferenciar regiones internas de la imagen de un grano se utiliza el concepto matemático de punto interior en regiones.

Los bordes en una imagen son necesarios, para poder aplicar diseñar una modelo de reconocimiento de regiones de interés en una imagen, por esta razón, se diferencian del resto de la imagen; para ello, se le da un valor específico a los bordes, mientras el resto de la imagen, conformada por puntos interiores y exteriores de la región de interés se asignan otros valores específicos como son el color negro y el azul respectivamente; esta técnica es conocida como binarización (Burger & Burge, 2009), que permite localizar fácilmente los bordes. A continuación se describen las etapas desarrolladas para la realización del proyecto.

3.1 DESCRIPCIÓN GENERAL DEL ALGORITMO

El objetivo de generar un algoritmo para caracterizar partículas es que, a pesar de las condiciones tan diferentes entre las muestras estudiadas (tipo de iluminación, características de la partícula, ausencia o presencia de ruido), éste sea capaz de segmentar correctamente las regiones de interés, para posteriormente realizar el procesamiento de las regiones y tomar decisiones. Para lograr esto, el algoritmo debe ser robusto al ruido y a los cambios en la iluminación presentes en imágenes reales. En esta tesis las partículas son los granos (partidos o enteros) presentes en la imagen digital. El algoritmo está estructurado en cinco etapas que son: Lectura de la

imagen, Selección de Umbrales, Filtrado de las zonas de interés, Segmentación y Extracción y análisis de las características. La siguiente figura nos ilustra la secuencia de los pasos en el algoritmo a implementar para el tratamiento de imagen.

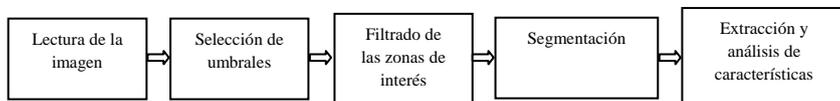


Figura 2. Diagrama en bloque del algoritmo

EN LA ETAPA 1

La imagen es representada de forma digital como un mapa de bits, lo cual permite convertir dicho mapa en una matriz de $M \times N$ elementos. Cada uno de los vértices en la matriz $I(i,j)$ (figura 2.), representa un píxel. Cada píxel tiene un valor numérico asociado, que representa el valor de intensidad o brillo del píxel dentro de la imagen.

	Color 1	Color 2	Color 3	Color 4		Color 5	Color 6
	Color 7	Color 8	Color 9	Color 10		Color 11	Color 12
j	Color 13	Color 14	Color 15	Color 16		Color 17	Color 18
	Color 19	Color 20	Color 21	Color 22		Color 23	Color 24
	Color 25	Color 26	Color 27	Color 28		Color 29	Color 30

Figura 3. Imagen como mapa de bits

Las matrices de las imágenes se caracterizan por ser de gran tamaño, puesto que el número de filas (M) y el de columnas (N) es del orden de 10^2 ó 10^3 , lo que implica sistematizar las operaciones con matrices, agilizando así los procesos matemáticos para el análisis de la imagen.

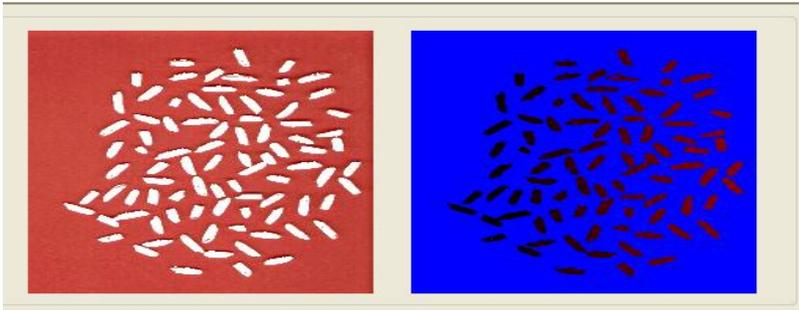


Figura 5. Proceso de segmentación de la imagen digital

Para poder contar cada uno de los granos o regiones, es necesario asignarle un valor diferente a cada región. Para lograr esto se recorrerá la matriz que contiene las regiones homogéneas, al encontrar un pixel homogéneo, es necesario etiquetarlo, pero es importante saber si es un punto interior o de frontera, para esto se hace el siguiente análisis a los pixeles de las regiones homogéneas.

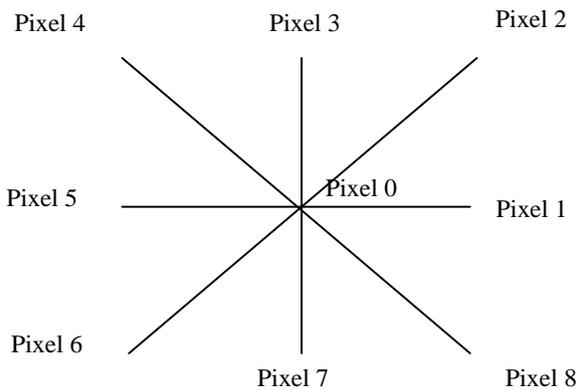


Figura 6. Representación molecular del análisis de los pixeles

Las condiciones en el análisis son:

- ✓ Si el pixel 0 o pixel que se analiza, y todos los pixeles a su alrededor (ver figura 5) tienen el valor homogéneo o valor segmentado (ya enumerado) entonces el pixel 0 es un punto interior.
- ✓ Si pixel 0 y alguno de los pixeles a su alrededor tienen el valor homogéneo o valor segmentado entonces pixel 0 es un punto de frontera.

Una vez etiquetados los puntos interiores y de frontera, es necesario analizar su conectividad para determinar la región, para esto se analiza la distancia entre los pixeles de cada región segmentada, lo cual facilita determinar el área y el perímetro de cada grano o región en pixeles. El área de una región está dada por $A = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} g(i, j)$ donde $g(i, j)$ es un punto interior o de frontera. Mientras que el perímetro de una región está dado por $P = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} g(i, j)$ donde $g(i, j)$ es un punto de frontera. La siguiente figura nos representa lo anteriormente explicado:

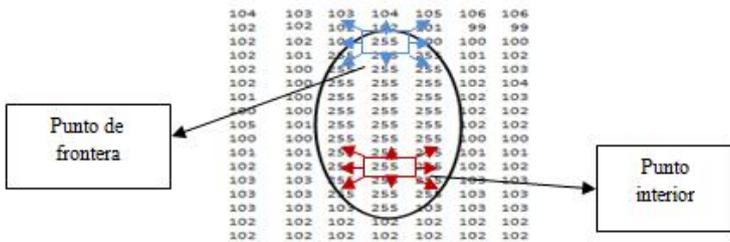


Figura 7. Representación puntos interiores y de frontera

Y en la etapa 5. Una vez segmentada la imagen es importante considerar la forma de describir los objetos que aparecen después de realizar el proceso de segmentación. Las características son usadas para describir un objeto o sus atributos. La extracción de las características es el paso a la clasificación. Las características serán extraídas por las propiedades topológicas y métricas de la región. Una vez aplicado lo anterior se aplicarán algunos métodos estadísticos con el objetivo de ordenar, clasificar e inferir sobre las

características del arroz estudiado. En el programa esta etapa se refleja en los siguientes ítems

➤ **SELECCIÓN DE LA MUESTRA:**

Para el cual se utiliza el **método de Bernoulli**¹. Que consiste en ordenar y asignar un valor $k=1, 2, 3, \dots, N$, a los elementos de una población, tomar un valor fijo π comprendido en el intervalo $(0, 1)$. Luego se asigna un valor aleatorio $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_N$ a cada uno de los elementos k , con $0 < \varepsilon_k < 1$. La selección o no selección, de los k th elementos es decidida por la siguiente regla: si $\varepsilon_k < \pi$, el elemento k es seleccionado sino no. Claramente, la probabilidad de selección es $P(\varepsilon_k < \pi) = \pi$. Este muestreo se aplica para escoger el número de sacos o bultos de arroz y luego al número de muestras que se desean sacar de cada uno de los bultos de arroz seleccionados por el método de Bernoulli.

➤ **ANÁLISIS MULTIVARIADO**²:

Se hace un análisis a las siguientes variables: área de granos partidos, área de granos enteros, color y peso de los granos. Para observar el comportamiento o participación de cada una de las variables capturadas.

4. RESULTADOS Y ANÁLISIS DE APLICAR LA TÉCNICA BASADA EN LAS MATEMÁTICAS Y EL TRATAMIENTO DE IMÁGENES

A continuación se evidencian los resultados y análisis al aplicar la nueva técnica basada en las matemáticas y tratamiento de imágenes para determinar la calidad del arroz.

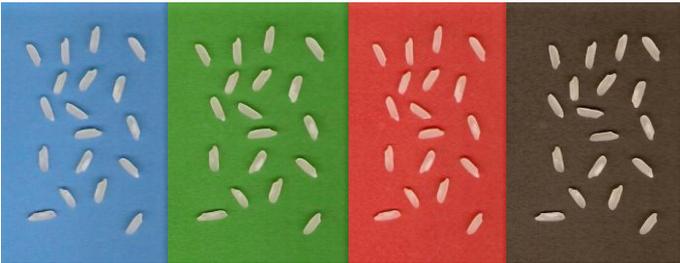
4.1 Caso de estudio 1

Se analizarán 24 granos de arroz, de los cuales hay 6 partidos y 18 completos. Se tomarán imágenes con los 18 granos enteros y con los 24 granos, pero con fondos de color diferente, y se analizarán con el programa desarrollado.

4.1.1 METODOLOGÍA PARA LA DIGITALIZACIÓN DE LA MUESTRA

Para obtener las imágenes de los granos de arroz, se utilizó un escáner, como resultado se obtuvieron las siguientes imágenes.

Imágenes de los 18 granos de arroz entero



Imágenes de los 24 granos de arroz, con 6 granos partidos

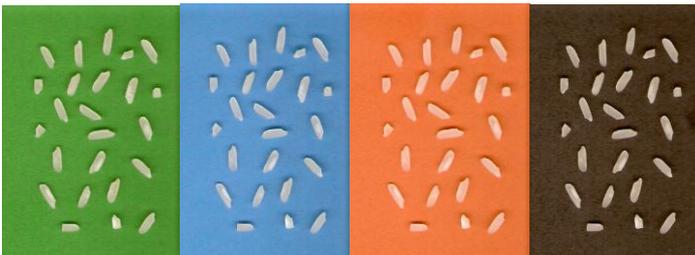


Figura 8. Imágenes de las muestra

4.1.2 RESULTADOS OBTENIDOS

Se obtuvieron excelentes resultados, como fueron: En el análisis de la imagen de las muestra de arroz se obtienen los siguientes logros: el código del color compuesto del grano, los códigos de cada uno de los RGB (red, green, blue) que son la composición del color compuesto en términos de la intensidad de los colores primarios de la luz, el número de granos observados por muestra, el tamaño de cada uno de los granos, el número de granos enteros y partidos así como el peso de la muestra.

En cuanto a los resultados estadísticos de estas muestras se obtiene: un análisis multivariado de las variables, área de granos enteros, área de granos partidos, color y peso. Con base en estas variables se calculan las matrices de covarianzas y correlación. Esta información facilita el cálculo de varianza total y la generalizada, con el objeto de calcular el peso de cada una de estas variables con respecto a la varianza del total, para así apreciar el papel juegan estas variables en el cálculo multivariado.

En cuanto al análisis estadístico de cada una de las muestras, se obtiene un reporte que contiene el número de muestras analizadas, así como el número de granos observados, el número de granos enteros, el porcentaje de granos enteros, el número de granos partidos, el porcentaje de granos partidos, el código del color compuesto y el peso de la muestra. Por último se hace la inferencia sobre los granos partidos en la población. Del cual se analiza el intervalo de confianza que contempla la proporción de granos partidos de la población de arroz de donde provienes dichas muestras. El grado de humedad de los granos se determina con base en el análisis que se hace de los códigos de los colores primarios de la luz o los RGB (red, green, blue).

En las figuras siguientes se muestran la forma como fueron procesadas las muestras de arroz analizadas haciendo uso del método por medio de una aplicación computacional programada con base en el método matemático explicado anteriormente.

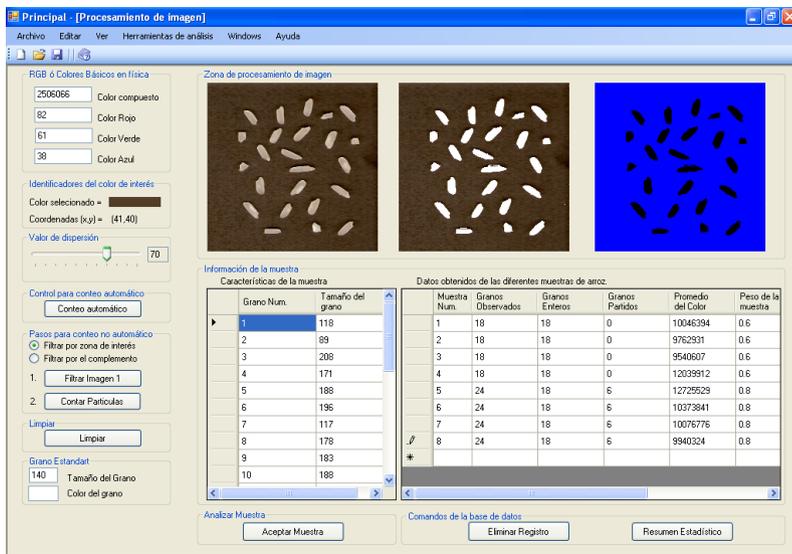


Figura 9. Imagen del resultado de las ocho muestras

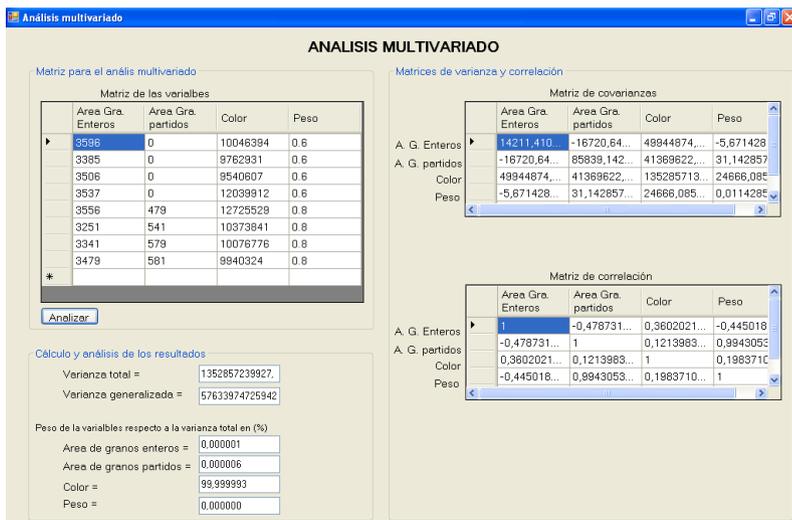


Figura 10. Análisis multivariado de las muestra

4.1.3 Análisis de los resultados

Con base en los resultados obtenidos se puede observar la precisión que tiene el programa en el tratamiento de imágenes, pues la información capturada de la imagen por la aplicación es la que se deseaba obtener. Es importante resaltar que en las muestras analizadas unas tenían 18 granos enteros y otras contenían 24 granos de los cuales había 18 buenos y 6 partidos, resultados que arrojó la aplicación al procesar las muestras con diferentes fondos de color. La cuestión de cambiar los colores de los fondos es un factor importante para medir la robustez de la herramienta computacional.

Una vez analizadas las muestras, se puede hacer un análisis multivariado, para observar el comportamiento de las variables que estuvieron implícita en el proceso, estas son: área de granos enteros, área de granos partidos, color y peso. Según el reporte multivariado de la aplicación los pesos respectivos de cada una de estas es: la variable con mayor peso es el color con el 99,999993%, en segundo lugar se encuentra el área de granos partidos con un 0,000006%, en tercer lugar está el área de granos enteros con un 0.000001% y por último está el peso (gr) quien no obtuvo ningún peso porcentual. Es importante resaltar que estos pesos porcentuales de participación se calculan respecto a la varianza total cuyo resultado es 1352857239927, valor que representa la traza de la matriz de covarianza. Mientras que el valor de varianza generalizada es valor del determinante de la matriz de covarianza (ver resultados en la figura anterior)

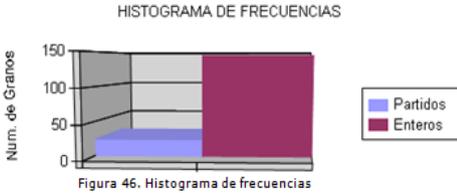
El formulario de análisis de las muestras, presenta un reporte general de los granos observados. En él se encuentra un resumen descriptivo e inferencial de las muestras analizadas. El estimador de proporción es el que utiliza la herramienta computacional para inferir sobre el porcentaje de granos partidos en la población. Se escogió este estimador porque cumple con las condiciones del mismo. En cuanto al nivel de confianza, el usuario lo puede escoger acorde a sus exigencias, pues el programa cuenta con esa opción. A continuación se presentan los resultados estadísticos obtenidos de la prueba practicada, con nivel de confianza del 95%.

REPORTE DEL ANALISIS DE LOS GRANOS DE ARROZ

Num de Muestras	Granos observados	Granos enteros	Granos partidos	Porcentaje de granos partidos	Color promedio	Peso de la muestra	Peso de la población Kgr
8	168	144	24	14,285714%	10563289	5,6	0

Tabla 2. Resumen descriptivo de las muestras

1. Analisis de la muestra



Proporción de granos enteros y partidos en la muestra

% de Granos enteros de la muestra	% de Granos partidos de la muestra
85,714286%	14,285714%

Tabla 3. Porcentaje de granos partidos

2. Inferencia sobre los granos partidos en la población

* Intervalo de confianza que indica la proporción de granos que partidos

$$0,089942 < \text{Prop de granos partidos} < 0,195772$$

Se estima que población estudiada contiene un porcentaje de granos partidos entre 8,9942% y el 19,5772%

* Los granos presentan humedad?: No

Los resultados anteriores muestran la precisión que tiene la nueva técnica basada en el análisis matemático y el tratamiento de la imagen para determinar la calidad del arroz. Además es importante resaltar que la aplicación implementada computacionalmente emplea un tiempo de 4 segundos para analizar cada una de las muestras.

5. CONCLUSIONES

En este trabajo se presentó una técnica basada en el análisis matemático y el tratamiento de imágenes para determinar el porcentaje de granos partidos en muestras de arroz. La cual está diseñada con un algoritmo que consta de cinco (5) etapas para el análisis y procesamiento de una imagen digital.

En la primera etapa del algoritmo, la imagen I, se le convierte en una matriz de MxN, cuyos elementos I(i, j) representan el valor del color compuesto de cada uno de los pixeles que conforman la imagen. En la segunda etapa, se identifica el umbral o color de la

zona de interés que se desea analizar, el umbral es fundamental para la construcción de los histogramas, el cual contiene el intervalo de códigos de colores que se desean filtrar, es decir:

umbral – dispersión < umbral < umbral + dispersión.

La tercera etapa, consiste en homogeneizar el color de los granos, los cuales son las regiones de interés, para lograr esto es necesario articular el umbral seleccionado con el histograma de la imagen, ya que este permite la cuantización o selección de los RGB que se desean filtrar de la matriz digital. En la cuarta etapa, una vez homogeneizadas las regiones, se aplican los conceptos de análisis matemático y topología como son: punto interior, frontera, conectividad y métricas, con el fin de etiquetar las zonas de interés para poder contar y medir el número de regiones o granos presentes en la imagen analizada.

En la quinta y última etapa, se analizan las variables capturadas de la imagen, las cuales son: color, perímetro y área de las regiones. Estas características son usadas para describir un objeto o sus atributos. La extracción de las características es el paso a la clasificación. Las características son extraídas por las propiedades topológicas y métricas de la región. Una vez aplicado lo anterior se aplican algunos métodos estadísticos con el objetivo de ordenar, clasificar e inferir sobre las características del arroz analizado.

Para determinar el rendimiento de la aplicación se analizaron varias muestras con el fin de observar y medir su precisión, en los resultados de dichas experiencias, se apreció que el programa fue muy preciso y rápido en el procesamiento de la información, lo cual se puede constatar visualmente en los pantallazos capturados de la técnica implementada en el programa computacional realizado en el lenguaje de Visual Basic. Con base en los resultados obtenidos se concluye que la técnica creada es eficaz para determinar el porcentaje de granos partidos en muestras de arroz, factor de mucha importancia para la clasificación y calidad del grano.

REFERENCIAS BIBLIOGRÁFICAS

- CROISIER, Bernard. 1998. The governance of external research: empirical test of some transaction-cost related factors. **R&D Management**. Nº 28: 289-298.
- DÍAZ MARTÍN, María Carmen. 1996. Factores determinantes de la innovación tecnológica para las empresas pequeñas. **Cuadernos de Estudios Empresariales**. Nº 6: 145-154.
- DORASZELSKI, Ulrich y JAUMANDREU, Jordi. 2013. R&D and Productivity: Estimating Endogenous Productivity. **Review of Economic Studies**. Nº 80: 1338 – 1383.
- GRILICHES, Zvi. 1980. R & D and the Productivity Slowdown. **American Economic Review**. Vol. 70. Nº 2: 343-348.
- HAGEDOORN, John. 2002. Inter-firm R&D partnerships: An overview of major trends and patterns since 1960. **Research Policy**. Nº 31: 477-492.
- MARITI, P. y SMILEY, R. 1983. Co-operative agreements and the organization of industry. **Journal of Industrial Economics**. Vol. 31, Nº 4: 437-451.
- MOWERY, David C. 1983. The relationship between intrafirm and contractual forms of industrial research in American manufacturing, 1900-1940. **Exploration in Economics History**. Nº 20: 351-374.
- NELSON, Richard. 2000. "National innovation systems", en: Acs, Z. (ed.), **Regional Innovation, Knowledge and Global Change**: London.
- PISANO, Gary P. 1989. Using equity participation to support exchange: evidence from the biotechnology industry. **Journal of Law, Economics, and Organization**, Vol. 5. Nº 1: 109-126.
- PISANO, Gary P. 1990. The R&D Boundaries of the Firm: An Empirical Analysis. **Administrative Science Quarterly**. Vol. 35, Nº 1: 153-176.

- SAKAKIBARA, Mariko. 1997. Heterogeneity of firm capabilities and co-operative research and development: an empirical examination of motives. **Strategic Management Journal**. N° 18: 143–164.
- VEUGELERS, Reinhilde. 1998. Collaboration in R&D: An Assessment of Theoretical and Empirical Findings. **De Economist**. N° 146: 419–443.
- WILLIAMSON, Oliver E. 1985. **The Economics Institutions of Capitalism**. Editorial Free Press. New York.
- WILLIAMSON, Oliver E. 1981. The Economics of Organization: The Transaction Cost Approach. **American Journal of Sociology**. N°. 87: 548-577.

Notas

¹SARNDAL, Carl-Erik, SWENSSON, Bengt y WRETMAN, Jan. 1992. **Model Assisted Survey Sampling**. Ed. Springer Verlag, Pág. 26, New York.

²DÍAZ, Luis (2002)



**UNIVERSIDAD
DEL ZULIA**

opción

Revista de Ciencias Humanas y Sociales

Año 33, N° 82, 2017

Esta revista fue editada en formato digital por el personal de la Oficina de Publicaciones Científicas de la Facultad Experimental de Ciencias,

Universidad del Zulia.
Maracaibo - Venezuela

www.luz.edu.ve

www.serbi.luz.edu.ve

produccioncientifica.luz.edu.ve