

UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS

FACULTAD DE INGENIERÍA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



**SOFTWARE AUTOMÁTICO DE RECONOCIMIENTO DE PATRONES PARA
LA DISCRIMINACIÓN DE GRANOS DE ARROZ PULIDO BASADO EN
PROCESAMIENTO DIGITAL DE IMÁGENES**

TESIS

PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS

LÍNEA DE INVESTIGACIÓN : COMPUTACIÓN GRÁFICA

AUTOR : Br. Joel Palomino Cáceres

ASESOR : Mgr. Cecilia Edith García Rivas Plata

PERÚ, 2014

DEDICATORIA

Con todo mi cariño y amor a las personas que hicieron todo en la vida para lograr mis sueños, por motivarme y darme la mano cuando sentía que el camino se terminaba. A ustedes papá y mamá por siempre mi corazón y mi agradecimiento.

Joel,

AGRADECIMIENTO

A mis padres Juan y Rosa por darme la vida.

A mis hermanos Jhon, Joe y Juan por su apoyo incondicional.

A mi asesor por guiarme en la culminación de esta investigación.

Muchas gracias.

**“La vida es solo una visión... un sueño, nada existe... solo espacio vacío
y tú, ...y tú no eres más que un pensamiento”**

TABLA DE CONTENIDO

	pág.
DEDICATORIA.....	ii
AGRADECIMIENTO	iii
TABLA DE CONTENIDO	v
LISTA DE FIGURAS	ix
LISTA DE TABLAS	x
RESUMEN	xi
ABSTRACT	xii
INTRODUCCIÓN	xiii
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA	1
1.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA	1
1.1.1. El arroz y su situación en el Perú y el mundo	1
1.1.2. Clasificación de granos	3
1.2. FORMULACIÓN DEL PROBLEMA	5
1.2.1. Problema general	5
1.2.2. Problemas específicos	5
1.3. OBJETIVOS DE LA INVESTIGACIÓN	6
1.3.1. Objetivo general	6
1.3.2. Objetivos específicos	6
1.4. JUSTIFICACIÓN.....	7
1.5. DELIMITACIÓN DE LA INVESTIGACIÓN	9
CAPÍTULO II: MARCO TEÓRICO.....	10
2.1. ANTECEDENTES DE LA INVESTIGACIÓN	10
2.1.1. Determinación de las características de aspecto de granos de arroz pulidos empleando procesamiento de imágenes	10
2.1.2. Sistema automático de diagnóstico de parásitos intestinales a través de imágenes digitales.....	11

2.1.3. Detección del Bacilo <i>Mycobacterium Tuberculosis</i> mediante reconocimiento de patrones.....	12
2.1.4. Análisis de imágenes microscópicas para la determinación de la cantidad y el tamaño de larvas de concha de abanico	13
2.1.5. Diseño y desarrollo de un algoritmo que permite estimar el tamaño de peces, aplicando visión por computadora, y propuesta realizar la selección adecuada de dichos peces	14
2.2. MARCO TEÓRICO	15
2.2.1. El arroz	15
2.2.2. Visión artificial	16
2.2.3. Imagen digital	17
2.2.4. Pixel	17
2.2.5. Profundidad de color	18
2.2.6. Clasificación de las imágenes	21
2.2.7. Procesamiento digital de imágenes	22
2.2.8. Detectores de borde.....	23
2.2.9. Operaciones morfológicas.....	26
2.2.10. Descriptores de forma	28
2.2.11. Segmentación de imágenes.....	30
2.2.12. Reconocimiento de patrones	34
CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN	36
3.1. TIPO DE INVESTIGACIÓN	36
3.2. MÉTODO DE INVESTIGACIÓN	36
3.3. ALCANCE DE LA INVESTIGACIÓN	36
3.4. DISEÑO DE INVESTIGACIÓN	37
3.5. POBLACIÓN Y MUESTRA DE ESTUDIO	38
3.5.1. Población	38
3.5.2. Muestra	38
3.6. SISTEMA DE HIPÓTESIS.....	38
3.6.1. Hipótesis principal	38
3.6.2. Hipótesis secundarias	39
3.7. SISTEMA DE VARIABLES.....	39
3.7.1. Variable independiente	39

3.7.2.	Variable dependiente	39
3.7.3.	Relación entre las variables	40
CAPÍTULO IV: DESARROLLO DE LA SOLUCIÓN		41
4.1.	PLAN DE DESARROLLO DEL SOFTWARE	41
4.1.1.	Beneficio	41
4.1.2.	Hardware y software empleado	42
4.2.	ESTUDIO DE FACTIBILIDAD	43
4.2.1.	Viabilidad técnica	43
4.2.2.	Viabilidad económica	44
4.2.3.	Viabilidad operativa.....	45
4.2.4.	Relación costo beneficio.....	46
4.3.	DESARROLLO	46
4.3.1.	Metodología de desarrollo	46
4.3.2.	Análisis de requerimientos	48
4.3.3.	Especificaciones de los casos de uso	49
4.3.4.	Diagrama de los casos de uso.....	52
4.4.	DESCRIPCIÓN DE LA SOLUCIÓN DESARROLLADA	53
4.4.1	Adquisición de la imagen.....	53
4.4.2.	Pre procesado de la imagen	55
4.4.3.	Segmentación de la imagen	59
4.4.4.	Extracción de características	59
4.4.5.	Discriminación de objetos	61
4.4.6.	Implementación del software	63
CAPÍTULO V: RESULTADOS.....		65
5.1.	PRESENTACIÓN DE RESULTADOS	65
5.1.1.	Resultados del experimento Nro. 01.....	65
5.1.2.	Resultados del experimento Nro. 02.....	65
5.1.3.	Resultados del experimento Nro. 03.....	66
5.2.	COMPROBACIÓN DE HIPÓTESIS.....	69
5.2.1.	Comprobación de las hipótesis secundarias	69
5.2.2.	Comprobación de la hipótesis principal	70

5.3. ANÁLISIS Y DISCUSIÓN DE RESULTADOS	71
5.3.1. Experimento Nro. 01	71
5.3.2. Experimento Nro. 02	71
5.3.1. Experimento Nro. 03	71
CAPÍTULO VI: CONCLUSIONES	73
CAPÍTULO VII: RECOMENDACIONES.....	74
BIBLIOGRAFÍA	75
ANEXOS	78
ANEXO 1: DISTRIBUCIÓN DE PESO DEL ARROZ.....	78
ANEXO 2: PARTES DE UN GRANO DE ARROZ	79
ANEXO 3: OPERACIONALIZACIÓN DE VARIABLES.	80
ANEXO 4: MATRIZ DE CONSISTENCIA.	81
ANEXO 5: MANUAL DE USUARIO	83
ANEXO 6: IMÁGENES RGB DE UN GRANO DE ARROZ CON EL ÁREA TIZOSA	86

LISTA DE FIGURAS

	pág.
Figura 1: Producción de granos de arroz a nivel mundial en millones de toneladas (2010-2013)..	3
Figura 2: Producción de granos de arroz a nivel mundial por continentes (2010-2013).....	3
Figura 3: Ampliación de una zona de una imagen donde se pueden apreciar los pixeles.	18
Figura 4: Representación de una imagen RGB.....	19
Figura 5: Representación de una en escala de grises.	20
Figura 6: Representación de una imagen binaria.	20
Figura 7: Imagen ampliada en un 200 % respecto al de su tamaño original.	21
Figura 8: Imagen ampliada a diferentes niveles respecto al de su tamaño original.....	22
Figura 9: Representación de un pixel (Z5) y sus ocho pixeles vecinos.....	24
Figura 10: Ilustración de la operación morfológica apertura.	27
Figura 11: Ilustración de la operación morfológica cierre.	28
Figura 12: Ilustración del valor umbral (U) en el histograma de la imagen.....	32
Figura 13: Proceso de reconocimiento de patrones.....	35
Figura 14: Diagrama de los casos de uso.	52
Figura 15: Diagrama de bloques de la solución planteada.	53
Figura 16: Imagen escaneada de granos de arroz pulido con fondo blanco.....	54
Figura 17: Imagen escaneada de granos de arroz pulido con fondo negro.	55
Figura 18: Imagen binarizada con umbral obtenido con el método de Otsu.	56
Figura 19: Imagen binarizada donde se eliminó el ruido por completo.	57
Figura 20: Imagen RGB donde se eliminó el ruido en el fondo (negro) por completo.	58
Figura 21: Imagen RGB donde se eliminó el ruido en el fondo (blanco) por completo.	58
Figura 22: Imagen de los granos segmentados y etiquetado.	59
Figura 23: Imagen de los granos con la distancia máxima.	60
Figura 24: Imagen de los granos con el área defectuosa en el centro.....	61
Figura 25: Imagen de los granos discriminados en función a su clase.....	62
Figura 26: Imagen de los granos discriminados en función a su tipo de defecto.	63
Figura 27: Interfaz gráfica del Software automático.....	64

LISTA DE TABLAS

	pág.
Tabla 1: Superficie sembrada en hectáreas a nivel nacional.	1
Tabla 2: Producción y superficie cosechada a nivel nacional por regiones.	2
Tabla 3: Especificación del actor Operario.	49
Tabla 4: Descripción del caso de uso LeerImagen.	49
Tabla 5: Descripción del caso de uso BinarizarImagen.	50
Tabla 6: Descripción del caso de uso DiscriminarTipo.	50
Tabla 7: Descripción del caso de uso DiscriminarClase.	51
Tabla 8: Descripción del caso de uso ReducirRuido.	51
Tabla 9: Descripción del caso de uso GrabarImagen.	52
Tabla 10: Clase de granos de arroz pilado.	61
Tabla 11: Tipos de granos tizosos de arroz pilado.	62
Tabla 12: Resultados del Experimento Nro. 01.	65
Tabla 13: Número de granos discriminados en función a su clase.	65
Tabla 14: Número de granos en función a su tipo de defecto.	66
Tabla 15: Observaciones tomadas con el software.	67
Tabla 16: Observaciones tomadas con el Vernier.	68

RESUMEN

El objetivo general de la investigación es implementar un sistema capaz de determinar el número total de granos de arroz pulido en una imagen digital y discriminarlos en función a su clase (tamaño largo, mediano o corto) y tipo de defecto (grano normal, tizoso total o tizoso parcial) de manera automática utilizando procesamiento digital de imágenes. La hipótesis con la que se hizo la operacionalización de las variables fue la aplicación del software automático basado en procesamiento digital de imágenes que reduce el tiempo y margen de error en el conteo total de granos de arroz pulido y en el proceso de discriminación por su clase y tipo de defecto.

En la metodología empleada según su alcance fue explicativa. Según el diseño de contrastación fue mixto, ya que en uno de los experimentos se utilizó granos de arroz pulido comerciales formados al azar, y en el otro se utilizó granos con características morfológicas específicas escogidos de manera deliberada.

Los resultados que se obtuvieron luego del experimento mostraron una reducción significativa del tiempo y del margen de error en el conteo de granos de arroz pulido así como en la discriminación, en función a su clase y tipo de defecto; todo esto gracias a las características morfológicas capturadas a través del escáner a los granos de arroz pulido.

Se concluye que la aplicación de procesamiento digital de imágenes reduce el ruido en las imágenes digitales. El algoritmo desarrollado resalta las características morfológicas de los granos y facilita la obtención de información para su posterior interpretación. Los resultados obtenidos por el software en la reducción del tiempo y margen de error muestran una ventaja significativa frente al método convencional.

Palabras claves: Procesamiento digital de imágenes, algoritmo, segmentación.

ABSTRACT

The overall objective of the research is to implement a system capable of determining the total number of grains of polished rice in a digital image and discriminate according to class (along short, medium o) and type of defect (normal grain, Total tizoso or partial) tizoso automatically using digital image processing. The hypothesis with which the operationalization of the variables did was apply automatic software based on digital image processing that reduces the time and margin of error in the total count of polished rice grains and in the process of discrimination class and type of defect.

In the methodology as its scope was explanatory. According contrasting design was mixed as in one experiment commercial polished rice grains randomly formed was used, and the other grain was used with specific morphological characteristics deliberately chosen.

The results that were obtained after the experiment showed a significant reduction of time and the margin of error in counting polished rice grains as well as discrimination based on their class and type of defect; all thanks to the morphological characteristics captured through the scanner polished rice grains.

It is concluded that the application of digital image processing reduces noise in digital images. The algorithm developed highlights the morphological characteristics of the grains and facilitates obtaining information for further interpretation. The results obtained by the software in reducing the margin of error time and show a significant advantage over the conventional method.

Keywords: Digital image processing, algorithm, segmentation.

INTRODUCCIÓN

El arroz es el segundo cereal más producido en el mundo. La producción en el Perú, no es la excepción, ya que es considerado como un alimento básico en nuestra cultura culinaria. Actualmente se observa una creciente demanda de arroz en nuestro territorio. El control de calidad es una parte importante dentro de la producción, porque es un filtro de seguridad que debe cumplir para asegurarse si el grano reúne las características de calidad. El método convencional de realizar esta labor es extraer muestras de los lotes de grano, para ser medidos y discriminados en función de sus características morfológicas. Para determinar su clase se hace uso de un Vernier y de uno en uno se determina la longitud de todos los elementos de la muestra, para determinar el tipo de defecto que tiene el grano se realiza a través de una selección manual simple.

Como toda actividad humana, el proceso de selección manual está expuesto a cometer errores, por ser una tarea repetitiva y tediosa que provoca un cansancio visual y manual, lo que incrementa todavía más el error.

Con la finalidad de solucionar este problema, en la presente investigación se diseñó e implementó el software automático de reconocimiento de patrones basado en el procesamiento digital de imágenes.

La implementación de este software permite determinar el número total de granos y discriminarlos en función a su clase y tipo de defecto de manera automática. De esta manera, se puede prescindir del operario calificado y de las subjetividades del mismo.

El trabajo está organizado de la siguiente manera:

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA DE ESTUDIO. En el que se exponen la fundamentación, descripción de la realidad problemática, formulación del problema, objetivos, justificación y delimitación de la investigación.

CAPÍTULO II: MARCO TEÓRICO. En esta parte exponemos los antecedentes de la investigación, el marco teórico teniendo en cuenta investigaciones importantes de autores nacionales e internacionales que constituyeron la base científica de la investigación.

CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN. Aquí se describen, la metodología de la investigación, tipo de investigación, método de investigación, alcances de la investigación, diseño de investigación, población y muestra de estudio, sistema de hipótesis y sistema de variables.

CAPÍTULO IV: DESARROLLO DE LA SOLUCIÓN. En esta parte explicamos el plan de desarrollo del software, el estudio de factibilidad, el desarrollo del software y una descripción de la solución desarrollada.

CAPÍTULO V: RESULTADOS. Aquí se analiza y se interpreta los datos obtenidos a través de los distintos experimentos.

CAPÍTULO VI: CONCLUSIONES. Aquí se muestran las conclusiones a las cuales se arribaron en base a los resultados obtenidos.

CAPÍTULO VII: RECOMENDACIONES. En esta parte se recomienda tener en cuenta algunos aspectos técnicos para extender los objetivos del presente trabajo.

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

1.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA

1.1.1. El arroz y su situación en el Perú y el mundo

En los últimos años, la producción nacional muestra un crecimiento favorable y muestra una evolución positiva a lo largo de los últimos 62 meses en casi todos los sectores, parte de este crecimiento se encuentra en el sector agropecuario (INEI, 2014). La producción de arroz es una de las actividades más importantes dentro de este sector y es el producto que más aporta al PBI Agropecuario y Agrícola. Se encuentra entre los principales cultivos del Perú, con una superficie total sembrada de 387,677 hectáreas (campaña agrícola 2011-2012) (ver tablas 1 y 2) y genera 162,300 empleos anuales permanentes, la mayor cantidad en este sector (Dirección General de Competitividad Agraria, 2013).

Superficie cultivada a nivel nacional

Campaña	Ago	Set	Oct	Nov	Dic	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago-Jul	Var %
2006-2007	30,229	23,172	28,594	28,970	38,981	39,530	59,580	22,892	11,829	15,983	18,517	25,060	343,334	
2007-2008	26,036	29,209	28,213	32,954	39,681	47,022	58,679	28,017	15,609	19,352	20,621	23,101	368,492	7.30%
2008-2009	37,450	29,019	29,178	39,078	51,137	67,472	44,934	29,777	13,330	18,161	21,332	25,267	406,134	10.20%
2009-2010	33,289	29,482	28,275	33,815	43,303	52,538	73,335	25,396	14,206	17,736	18,679	25,787	395,839	-2.50%
2010-2011	33,529	26,459	28,150	34,646	39,203	34,155	54,976	38,190	16,966	19,268	19,539	22,184	267,262	-7.20%
2011-2013	30,878	37,764	25,317	27,254	41,533	40,955	82,494	26,507	15,242	17,774	16,453	25,510	387,677	5.60%

Tabla 1: Superficie sembrada en hectáreas a nivel nacional.

Fuente: MINAG-OEEE.

Elaboración: MINAG-DGCA-DIA.

Producción y superficie cosechada a nivel nacional

Regiones	Superficie Cosechada(ha)			Producción(t)			
	2010	2011	Var %	2010	2011	Var %	Particip %
San Martín	76,244	78,425	2.90%	500,284	521,838	4.3%	19.9%
Piura	55,358	44,175	-20.20%	499,845	383,315	-23.3%	14.6%
La Libertad	32,351	31,276	3.30%	304,003	322,520	6.1%	12.3%
Lambayeque	52,899	38,186	-27.80%	407,735	282,166	-30.8%	10.8%
Amazonas	43,741	37,822	-13.50%	322,878	282,547	-12.5%	10.8%
Arequipa	18,455	19,292	4.50%	245,950	257,559	4.7%	9.8%
Cajamarca	28,843	28,044	-2.80%	217,976	216,413	-0.7%	8.2%
Tumbes	15,324	16,365	6.80%	128,642	141,607	10.1%	5.4%
Loreto	31,171	34,431	10.50%	85,918	103,403	20.4%	3.9%
Resto del País	34,274	31,587	-7.80%	118,142	112,659	-4.6%	4.3%
Total	388,660	359,603	-7.50%	2,833,383	2,626,038	-7.30%	100%

Tabla 2: Producción y superficie cosechada a nivel nacional por regiones.

Fuente: MINAG-OEEE.

Elaboración: MINAG-DGCA-DIA.

El arroz es el segundo cereal más producido en el mundo (ProdSTAT, 2013). Este se produce con muchos otros propósitos distintos al del consumo humano como la producción de bebidas y el consumo animal. La producción mundial va en aumento y a través de los últimos años se puede observar (ver figura 1) un crecimiento continuo. El continente con mayor producción de granos de arroz es el Asiático con un poco más del 90 % de la producción mundial (ver figura 2). Se considera un alimento básico en la cocina oriental.

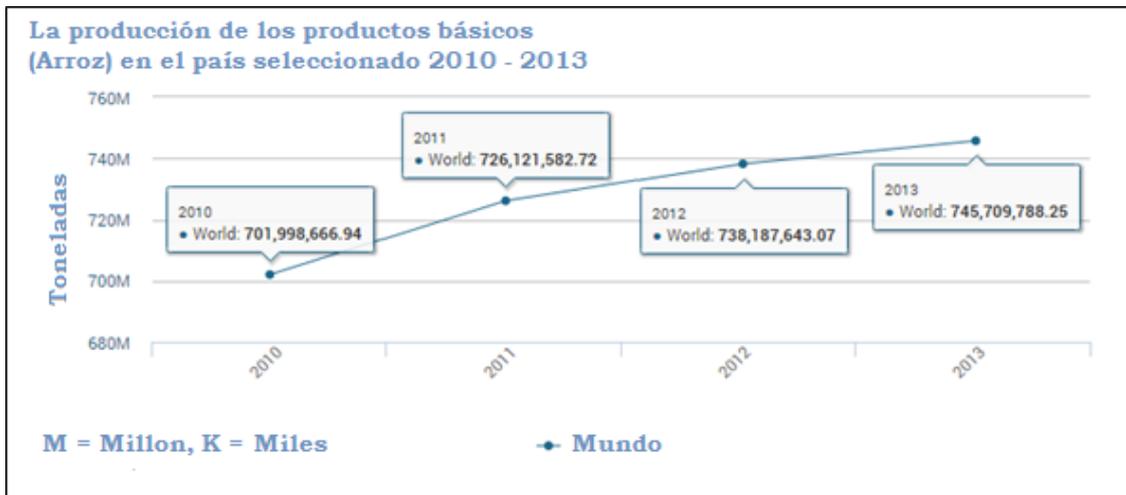


Figura 1: Producción de granos de arroz a nivel mundial en millones de toneladas (2010-2013).

Fuente: Organización de las Naciones Unidas para la Alimentación y Agricultura.

Elaboración: <http://faostat3.fao.org/>

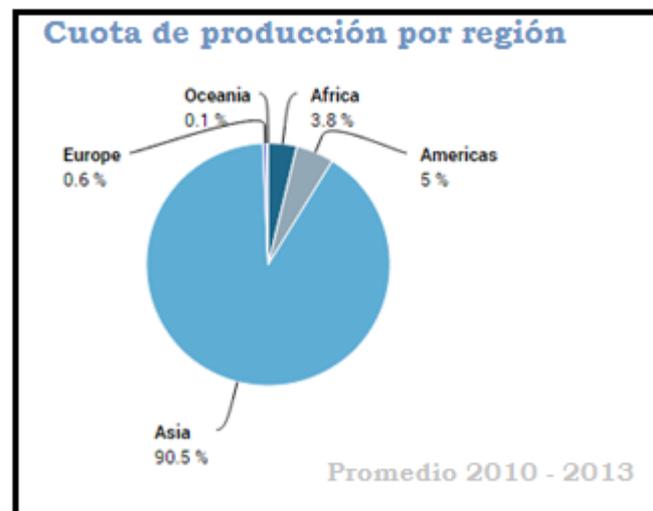


Figura 2: Producción de granos de arroz a nivel mundial por continentes (2010-2013).

Fuente: Organización de las Naciones Unidas para la Alimentación y Agricultura.

Elaboración: <http://faostat3.fao.org/>

1.1.2. Clasificación de granos

El arroz, al ser cultivado bajo condiciones y sistemas de producción diferentes, presenta variaciones en sus características en relación al tamaño, consistencia

al cocinar, cualidades aromáticas y color de grano, desde castaño, rojo, púrpura, negro hasta el blanco tradicional (Pérez & María, 2009).

Para obtener un producto en condiciones de ser consumido, el grano de arroz debe ser descascarado y pulido. Luego de este proceso se puede observar las áreas translúcidas del grano que son debidas a una disposición de los gránulos de almidón compacta y uniforme. Cuando la compactación no es homogénea existe una pérdida de la estructura cristalina y presenta formaciones de espacios de aire, se crean áreas opacas que originan la característica de Panza Blanca o Yesoso (Olmos, 2014).

El nombre “panza blanca” es conocido en el Perú como “Grano Tizoso parcial”, proviene de una mancha opaca blanca en la zona ventral del grano, que es menos frecuente en la parte dorsal.

Los granos “yesosos” también llamados “Granos Tizoso Totales”, presentan por más de la mitad de la extensión del grano la apariencia de blanco opaco.

La presencia de estos defectos tiende a aumentar el porcentaje de granos quebrados por ser más frágiles que los normales y afectan la calidad del producto, este defecto se asocia en un porcentaje alto a la heterogeneidad en el cultivo (diferentes grados de madurez), alto porcentaje de humedad en la cosecha o condiciones climáticas anormales como bajas temperaturas durante la madurez.

La calidad del grano de arroz está relacionada de entre otras características al tamaño y a la apariencia que tienen, es por eso que una de las etapas más importantes en la producción de arroz es el control de calidad, en donde se extraen muestras de los lotes para determinar el tamaño de los granos y la cantidad de granos defectuosos.

Un operario con la capacitación adecuada realiza el control de calidad, hace un conteo total para determinar el número de granos de la muestra, y con ayuda de un vernier mide cada grano para hallar el número total de granos largos, medianos y cortos. Del mismo modo identifica por medio de un método simple de separación si los granos poseen algún defecto como tizoso total, tizoso parcial o grano normal (Pérez & María, 2009).

Este método convencional demanda demasiado tiempo, y por tratarse de una determinación desarrollada por seres humanos está sujeta a errores como cualquier tipo de actividad humana. A eso se le suma lo tedioso y repetitivo de la tarea, lo que incrementa las posibilidades de cometer errores de distintos tipos, todo esto conlleva a tener resultados que no se ajustan a la realidad.

En una investigación realizada por (Pérez & María, 2009) en Aragua, Venezuela, indican que para la estimación del tamaño de granos de arroz hacen uso de un vernier electrónico, y para el conteo de granos con algún defecto usan un método simple de separación.

1.2. FORMULACIÓN DEL PROBLEMA

1.2.1. Problema general

¿Cuál es el efecto al aplicar un sistema automático basado en procesamiento digital de imágenes en la determinación del número total de granos de arroz pulido en una imagen digital y su discriminación en función a su clase y tipo de defecto?

1.2.2. Problemas específicos

Problema específico 1

¿Cómo el Software automático basado en procesamiento digital de imágenes afecta al tiempo en el conteo de los granos de arroz pulido?

Problema específico 2

¿Cómo el Software automático basado en procesamiento digital de imágenes afecta al tiempo en la discriminación de los granos de arroz pulido según su clase?

Problema específico 3

¿Cómo el Software automático basado en procesamiento digital de imágenes afecta al tiempo en la discriminación de los granos de arroz pulido según su tipo de defecto?

Problema específico 4

¿De qué manera el Software automático basado en procesamiento digital de imágenes se relaciona con margen de error en la medición de los granos de arroz pulido?

1.3. OBJETIVOS DE LA INVESTIGACIÓN

1.3.1. Objetivo general

Implementar un sistema capaz de determinar el número total de granos de arroz pulido en una imagen digital y discriminarlos en función a su clase y tipo de defecto, de manera automática, utilizando procesamiento digital de imágenes.

1.3.2. Objetivos específicos

Objetivo específico 1

Implementar un algoritmo que nos permita determinar el número total de granos en una imagen digital, basándose en el procesamiento digital de imágenes.

Objetivo específico 2

Diseñar un algoritmo que pueda discriminar los granos arroz pulido en función a su clase (tamaño largo, mediano o corto) basándose en el procesamiento digital de imágenes.

Objetivo específico 3

Diseñar un algoritmo que pueda discriminar los granos arroz pulido en función a su tipo de defecto (grano normal, tizoso total o tizoso parcial) basándose en el procesamiento digital de imágenes.

Objetivo específico 4

Diseñar un algoritmo que permita reducir el ruido en las imágenes digitales y resaltar las características de los granos para una mejor interpretación de datos.

Objetivo específico 5

Implementar en el software en MatLab R2010a los algoritmos para determinar el número total de granos, con el fin de realizar la discriminación de los granos de arroz pulido en función a su clase y tipo de defecto, que permita reducir el ruido en las imágenes digitales, resaltar las características de los granos y reducir el margen de error en la medición de longitud.

1.4. JUSTIFICACIÓN

Actualmente para determinar el número total de granos de arroz pulido, su distribución porcentual en función a su clase y tipo de defecto en una muestra, se usa un método manual que consiste en un proceso repetitivo y tedioso para el personal, ya que se mide con un vernier el tamaño de todos los granos para determinar su tamaño y su posterior clasificación, del mismo modo a través de un método simple de separación determinar si los granos tienen algún defecto (tizoso total o parcial) (Pérez & María, 2009). Este proceso como toda actividad humana está expuesta a cometer errores, a esto se le suma el cansancio visual provocado por la cantidad de granos que debe medir incrementando aún más la probabilidad de cometer errores en este proceso.

Estas razones son las que motivan diseñar e implementar un software automático con una aplicación concreta que determina el número total de granos de arroz pulido en una imagen digital, los discrimina en función a su clase (grano largo, mediano o corto) y tipo de defecto (grano normal, tizoso total o tizoso parcial). El mismo que no dependerá de la experiencia ni de la capacitación de algún personal. Por su facilidad de uso y amigable diseño.

Reduciremos el tiempo y el margen de error en las mediciones en comparación con el método convencional que por ser un método manual demanda demasiado tiempo y como toda actividad humana está sujeto a una probabilidad de error.

A través del procesamiento digital de imágenes se reducirá el ruido en las mismas, resaltar características relevantes, realizar conteos, determinar características morfológicas y facilitar la interpretación de los datos.

El software desarrollado ofrecerá apoyo a todas las personas que les resulta importante tener conocimiento sobre las características morfológicas de los granos de arroz pulido, como productores nacionales que no cuentan con avanzada tecnología e institutos de investigación que realizan esta caracterización morfológica de manera manual, ya que optimiza el proceso de discriminación y caracterización de estos granos, gracias a la información obtenida se podrá tomar decisiones adecuadas en las siguientes situaciones (SEDESOL, 2014):

Catalogar el grano: Se refiere a la identificación del grado de calidad del grano, considerando la Norma Técnica Peruana como referencia para posteriormente poder catalogarlo de acuerdo al rango al que pertenece.

Determinar el valor económico del grano: Se refiere al cálculo del valor económico en base a los parámetros de calidad que tiene una partida o lote de granos, para lo cual se deberá considerar características como: el porcentaje de humedad, el porcentaje de impurezas, granos dañados, granos defectuosos, etc.

Definir su uso o destino final: El conocer las características del grano, nos permite definir su uso final para el que puede ser apto como: el consumo humano, consumo animal, consumo industrial o en el último de los casos, para su destrucción.

Adoptar mejores medidas para su almacenamiento: Las condiciones de calidad del grano permiten definir la forma en la que es factible llevar a cabo su almacenamiento, así como las medidas preventivas que se deben tomar para evitar la pérdida de la calidad durante el periodo en que se encuentre en bodega.

Definir acciones necesarias para la conservación del grano: Los granos están expuestos a procesos de deterioro durante su almacenamiento,

por lo que es importante conocer su calidad evolutiva y en su caso, realizar acciones para evitar al máximo la pérdida de calidad.

1.5. DELIMITACIÓN DE LA INVESTIGACIÓN

La calidad del grano de arroz, *Oryza Sativa* L., considera aspectos molineros, culinarios, de apariencia y nutricionales (Pérez & María, 2009). En el presente trabajo de investigación se realizó el diseño e implementación de un sistema automático que permite obtener resultados a través de las características morfológicas capturadas en una imagen digital como: el número total de granos de arroz pulido, discriminar los granos en función a su clase (tamaño largo, mediano o corto) y tipo de defecto (grano tizoso total o parcial). Mostrar las imágenes digitales previamente procesadas con la reducción del ruido, y con el perímetro marcado (color rojo, amarillo o verde) en función a su discriminación.

Para determinar la calidad de los granos de arroz pulido hace falta hacer exámenes especializados y determinar otros tipos de características (ej. olor, humedad, densidad del grano, etc.), que no se obtienen a través de sus características morfológicas.

Las características morfológicas de un grano (clase y tipo de defecto), son sólo una parte de los indicadores de calidad.

Las imágenes digitales obtenidas para el experimento fueron capturadas a través de un escáner corriente de escritorio a 300 ppp. Los granos deben cumplir la única condición de no estar en contacto entre sí. El experimento fue realizado con muestras de granos de arroz pulido comercial de calidad extra.

CAPÍTULO II: MARCO TEÓRICO

2.1. ANTECEDENTES DE LA INVESTIGACIÓN

2.1.1. Determinación de las características de aspecto de granos de arroz pulidos empleando procesamiento de imágenes

Los investigadores (Sampallo, Acosta, Cleva, & González, 2013) implementaron una metodología basada en procesamiento digital de imágenes para determinar las características de aspecto de arroz. Emplearon el análisis estadístico de distribución de niveles de gris de los píxeles de la sección proyectada de cada grano en la imagen y se identificó en forma automática el tipo de grano: normal, panza blanca o yesoso.

El método empleado fue objetivo, rápido en la que se procesaron imágenes a color de muestras de 400 a 600 granos según la variedad. Los resultados de las muestras evaluadas fueron de dos tipos: a) Global: porcentajes de cada tipo sobre el total de granos de la muestra. b) Individual: describe por histograma el aspecto de cada grano. Además, las imágenes de la muestra se almacenaron en forma digital, lo que permite repetir las determinaciones sin pérdida de información.

Esta metodología desarrollada realiza una discriminación de los granos de arroz pilado por sus defectos en base a las especificaciones de calidad de los granos de arroz pulido definidos en el Código Alimentario Argentino (MINAGRI), en las normas de calidad de arroz envasados que están publicadas en el Boletín Oficial de la República Argentina (MINAGRI) y el Reglamento Técnico del MERCOSUR para Establecimiento de Identidad y de Calidad de Arroz Elaborado (MERCOSUR), donde identifican la calidad del grano de arroz según el largo del grano, la relación largo/ancho y su aspecto.

Los documentos mencionados también definen los granos defectuosos como son los granos panza blanca y yesosa. La denominación “panza blanca” proviene de una mancha opaca blanca en la zona ventral del grano. Este tipo de

granos degrada el aspecto del conjunto de granos translucidos y tiende a aumentar el porcentaje de granos quebrados por ser más frágil que los normales.

La metodología desarrollada permite la determinación del aspecto de granos de arroz pulido, empleando un software de visión artificial a través de un sistema experto que permite clasificar en forma rápida y eficiente, evitando subjetividades del ser humano. Los sistemas expertos son una rama de la inteligencia artificial que emulan el razonamiento de un experto en un dominio concreto, y en ocasiones son usados por estos. Con los sistemas expertos se busca una mejor calidad y rapidez en las respuestas, dando así lugar a una mejora de la productividad del propio experto al usar este tipo de sistemas informáticos. En ésta investigación usaron un clasificador de tipo neuro-reticular, es decir redes que se ‘entrenan’ para dar una cierta respuesta ante determinados valores de ingreso.

Los equipos y dispositivos empleados fueron: una computadora personal, un dispositivo de adquisición de imágenes y el software para realizar el procesamiento digital. Específicamente, se empleó un escáner de mesa para la adquisición de imágenes que permitió mantener constante las condiciones de iluminación de las diferentes muestras, las que fueron extraídas de paquetes comerciales de 1kg, cada muestra de 10 gr y se distribuyeron sobre el escáner, con la condición de que los granos no tengan contacto entre sí.

2.1.2. Sistema automático de diagnóstico de parásitos intestinales a través de imágenes digitales

En esta investigación (Beltran, 2011) se desarrolló un modelo computacional que permite de forma automática la detección, caracterización y diagnóstico de la especie de huevos de parásito que se encuentra en una muestra fecal, a través del reconocimiento de formas a partir de imágenes digitales.

Para la detección de los huevos de helmintos, aplicó la técnica de regiones salientes con el fin de permitir la segmentación de imágenes; utilizó, además, la técnica *multithresholding-kmeans* para la diferenciación entre los huevos y las impurezas de las muestras. El reconocimiento de patrones se basó en

características geométricas, morfológicas y de textura, habiendo conseguido un vector que contiene solo 8 características.

La caracterización de los objetos de interés comprende la extracción de características geométricas, morfológicas y de textura, habiendo conseguido conformar un vector que contiene solo 8 características que lograban describir correctamente a los 8 distintos tipos de huevos de Helmintos.

En el proceso de clasificación, se aplicó técnicas de aprendizaje supervisado, habiendo experimentado con clasificadores bayesianos (obteniendo un 93 % de acierto) y con Support Vector Machine (94 % de acierto). El software fue implementado en una plataforma web que permite realizar diagnósticos en tiempo real y obteniendo el resultado de manera automática.

Los resultados obtenidos llegaron a un 92 % de asertividad, una tasa muy alta. En cambio el promedio de asertividad de un experto humano es del 70 %.

Los equipos y dispositivos empleados fueron: una computadora, un microscopio y la creación de una interfaz que permitiera la comunicación entre ambos.

2.1.3. Detección del Bacilo *Mycobacterium Tuberculosis* mediante reconocimiento de patrones

(Romo, Ramírez, & Valdivieso, 2007), presentan una forma de automatizar el proceso de diagnóstico de la tuberculosis mediante un algoritmo basado en wavelets de reconocimiento de patrones morfológico y estadístico a través del procesamiento digital de imágenes, que permite aislar y cuantificar los bacilos en imágenes fotográficas de muestras de esputo que previamente fueron sometidas a un procesamiento digital para eliminar el ruido y resaltar características importantes. Todo esto para facilitar la interpretación de información por parte del software. Con esto se consigue emitir un diagnóstico automático y rápido del grado de infección o severidad de la infección con un alto grado de confiabilidad.

El diagnóstico de la enfermedad activa se realiza mediante la identificación y conteo manual de la población de los *bacilos de Koch* en el esputo, esto es un procedimiento ortodoxo que demanda tiempo y está sujeto a errores humanos.

El proceso empleado para obtener el diagnóstico de la infección de la tuberculosis considera los siguientes pasos:

- Toma de muestras de esputo del paciente.
- Pigmentación de las muestras para facilitar la identificación de los bacilos fotografiado digital de las muestras (en cualquier formato:*.tif, *.jpg, etc.).
- Sometimiento al software de diagnóstico automático desarrollado, obtención y análisis del diagnóstico y toma de decisiones.

Como resultados de este software en la detección y cuantificación del *bacilo de Koch* en imágenes digitales tomadas sobre muestras de un paciente infectado, se pudo concluir que: se llegó a un nivel de confiabilidad que oscila entre los 90 % y 95 %.

Cabe resaltar que el nivel de confiabilidad puede verse afectado por las manchas oscuras en el color del fondo, esto debido a que la pigmentación no se realizó de manera uniforme. Para solucionar este inconveniente, es necesario un ajuste en el nivel del umbral de color.

2.1.4. Análisis de imágenes microscópicas para la determinación de la cantidad y el tamaño de larvas de concha de abanico

Determinar el número y tamaño de larvas de conchas de abanico concentradas en los tanques de producción es una tarea ardua. El investigador (López, 2009) implementa una alternativa para optimizar este proceso: captura las imágenes a través de una cámara analógica y se conecta al computador a través de una tarjeta de adquisición de video.

Una vez que la imagen digital es capturada, se procede a realizar el recorte de la zona de interés de 1x1 mm marcada por los límites de cámara Sedgewick Rafter, se convierte la imagen a escala de grises para poder aplicar el filtro Sobel y pueda resaltar los bordes, seguidamente la imagen es binarizada usando un método de umbralización adaptativa.

Es ahí en donde se etiqueta a todos los elementos (larvas) de la imagen para ser medidos y contados. Cabe resaltar que el área promedio de la larva se relaciona

con el periodo de la señal del histograma. Usa la transformada rápida de Fourier para hallar la frecuencia del histograma, de manera experimental, se determinó que el área promedio de una larva es la inversa de la frecuencia multiplicado por la longitud del segmento considerado en el histograma.

El cálculo del diámetro se realiza aproximando cada larva al círculo más pequeño que lo pueda encerrar.

Los resultados muestran una exactitud de 95 % un valor significativo comparándolo con el método manual. El método es simple y efectivo que permitirá realizar un seguimiento continuo durante las etapas de crecimiento de las larvas.

Los equipos utilizados fueron: un microscopio MOTIC, modelo B3-223ASC, una cámara Sedgewick Rafter (lámina rectangular de vidrio con divisiones de 1x1 mm), una cámara de video digital Firewire Imaging Source DFK 21F04, una tarjeta PCMCIA, una laptop Pentium Core2duo, software de captura de imágenes, software MatLab.

2.1.5. Diseño y desarrollo de un algoritmo que permite estimar el tamaño de peces, aplicando visión por computadora, y propuesta realizar la selección adecuada de dichos peces

La tecnología usada en la acuicultura es precaria o no es adecuada para ser utilizada, esto ocasiona que los peces obtenidos no sean de buena calidad y se tenga una alta mortalidad en su producción. Esta es la razón que lleva a (Orellana, 2008) desarrollar un método más sofisticado, que permita obtener la longitud de los peces, sin que haya un contacto físico entre los peces y el hombre. Usando técnicas de visión artificial.

El alto nivel de mortandad de los peces en parte se debe a la mala manipulación en el proceso de selección. La técnica de medición de los peces es básicamente atraparlos sacarlos y medirlos.

El proceso de medición automatizado empieza con la captura de la imagen al pez, se procede a eliminar el ruido de la imagen con ayuda del software MatLab versión 7.1.0.246 que nos brinda altas prestaciones en el tratamiento de

imágenes y señales. Se seleccionó uno de los planos RGB en el que el pez muestra sus características de mejor manera para hacer el pre-procesamiento, en su caso el plano rojo.

Se eliminó las altas frecuencias que presentan los bordes usando el filtro pasa bajos, se procedió a binarizar la imagen con el método de Otsu porque mostró mejores resultados que las técnicas estadísticas, luego procedió a segmentar la imagen para detectar los bordes para su posterior medición.

Los resultados arrojan un nivel de confianza de 90 % lo que indica que se encuentra dentro de rango permitido para hacer una selección correcta de los peces, este algoritmo implementado es versátil ya que puede ser usado en distintos tipos de peces.

Los materiales utilizados fueron: una cámara de tipo CCD con código AVC 527 LN/F40 fotográfica que capture imágenes a una resolución de 320x240 como mínimo, tarjeta de adquisición de datos modelo CP-1400AS-(T), cable de conexión tipo RCA y una computadora con características básicas.

2.2. MARCO TEÓRICO

2.2.1. El arroz

El arroz es la semilla de la planta *Oryza sativa*. Se trata de un cereal considerado alimento básico en muchas culturas culinarias (en especial la cocina asiática), así como en algunas partes de América Latina.

Contiene una relativa pequeña cantidad de proteínas (en comparación con otros cereales), pues el contenido de gluten ronda el 7 % del peso, comparado con el 12 % de los trigos de bajo contenido de proteína. No obstante, el arroz posee más lisina que el trigo, el maíz y el sorgo (Kiple & Conee, 2000).

El arroz es el segundo cereal de mayor consumo en el mundo. la producción está geográficamente concentrada y más del 90 % proviene de Asia. Tan solo siete países asiáticos (China, India, Indonesia, Bangladesh, Vietnam, Myanmar y Tailandia) producen y consumen el 80 % de arroz del mundo (ProdSTAT, 2013).

La producción mundial de arroz con cáscara en el 2005 fue de 628 millones de toneladas. China es el mayor productor con 28 % del volumen total seguido de la India con 22 %, Indonesia con 9 %, Bangladesh y Vietnam con 6 %, Tailandia con 5 %, Myanmar con 4 % y otros países representan el 20 %.

Las variedades de arroz cultivadas han ido variando en los últimos años, mediante una gradual renovación de las más antiguas, en función de las mejores características, provocando la desaparición de determinadas variedades; pues las nuevas ofrecen mejores rendimientos, una mayor resistencia a enfermedades, mejor calidad de granos o una mayor producción. Los programas de mejora genética se basan en la producción de plantas de arroz *di haploides*, mediante el cultivo de anteras de plantas obtenidas a partir de cruzamientos previos (MINAG, 2009).

La distribución del peso del arroz pardo es la siguiente: pericarpio, 1-2 %; 5 % aleurona, nucela y cubierta seminal 4-6 %; germen 1 %; escutelo 2 %; y endospermo 90-91 %. (JULIANO, 1994).

El almidón forma el endospermo y se localiza en el arroz pulido en un 90 %. Está compuesto de 7 a 33 % de amilosa y el resto es amilopectina. La amilopectina es la responsable de la estructura del gránulo de almidón, el cual consiste de áreas cristalinas (cristales, micelas) y no cristalinas (amorfas, fase gel), arregladas en capas concéntricas (Sandoval, Rodríguez, & Fernández, 2005).

2.2.2. Visión artificial

Conocida como "Visión por Computadora", es un campo de la "Inteligencia Artificial", conjunto de todas aquellas técnicas y modelos que nos permiten la adquisición, procesamiento, análisis y explicación de cualquier tipo de información espacial del mundo real obtenida a través de imágenes digitales (EcuRed). El propósito de la visión artificial es programar un computador que "entienda" una escena o características de una imagen.

Con la visión artificial se pueden:

- Automatizar tareas repetitivas de inspección realizadas por operadores.

- Realizar controles de calidad de productos que no era posible verificar por métodos tradicionales.
- Realizar inspecciones de objetos sin contacto físico.
- Realizar la inspección del 100 % de la producción (calidad total) a gran velocidad.
- Reducir el tiempo de ciclo en procesos automatizados.
- Realizar inspecciones en procesos donde existe diversidad de piezas con cambios frecuentes de producción.

2.2.3. Imagen digital

Una imagen puede ser definida como una función de dos dimensiones $f(x, y)$, donde x e y son las coordenadas espaciales, y la amplitud de f de cualquier par de coordenadas es llamada la intensidad o nivel de gris de la imagen en ese punto. Cuando x, y y el valor de amplitud de f son finitos y discretos, la imagen es llamada imagen digital y cada punto de esta imagen es conocido como píxel (González & Woods, 2008). De esta manera, una imagen digital puede ser representada como un arreglo matricial de $N \times M$ píxeles.

2.2.4. Pixel

Un píxel o pixel, plural píxeles (acrónimo del inglés picture element, 'elemento de imagen'), es la menor unidad homogénea en color que forma parte de una imagen digital (RAE).

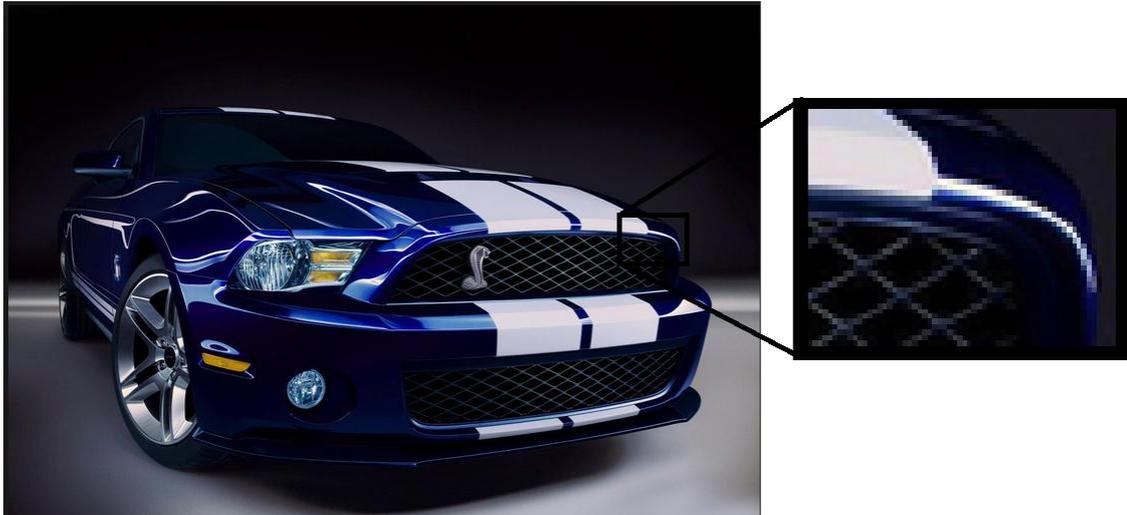


Figura 3: Ampliación de una zona de una imagen donde se pueden apreciar los píxeles.
Imagen: Obtenida de (Mustang).
Elaboración: Propia.

Ampliando una imagen digital (zoom) en la pantalla de una computadora, pueden observarse los píxeles que componen la imagen (ver figura 3). Los píxeles son los puntos de color que componen una imagen y define solamente un color.

(López, 2009) Indica las características más importantes de un pixel:

- Tienen una forma cuadrada
- Una posición relativa al resto de píxeles de un mapa de bits
- Una profundidad de color (capacidad para almacenar color), que se expresa en bits

2.2.5. Profundidad de color

Se llama profundidad de color al término que describe el número o la cantidad de bits utilizados para representar el color de un solo píxel (un bit es un dígito binario, que tiene un valor de 0 o 1). Este concepto de profundidad también se conoce como bits por píxel (bpp), es decir, la cantidad de color que hay en cada píxel, y se ocupa como unidad de medida sobre todo cuando se especifica junto con el número de bits utilizados. Mientras mayor es la profundidad de color se da

una gama más amplia de distintos colores (Calvo). Las imágenes se caracterizan por la cantidad de colores que contienen:

2.2.5.1. Imagen a colores RGB

RGB es el acrónimo de los colores en inglés: red, green, blue; en español: rojo, verde y azul. Es la composición del color en términos de la intensidad de los colores primarios de la luz. Las imágenes RGB están compuestas por tres capas donde cada una representa un color (ver figura 4). Tienen 256 tonalidades que van desde el [0 al 255], esto significa que cada pixel necesita 3 bytes o 24 bits.

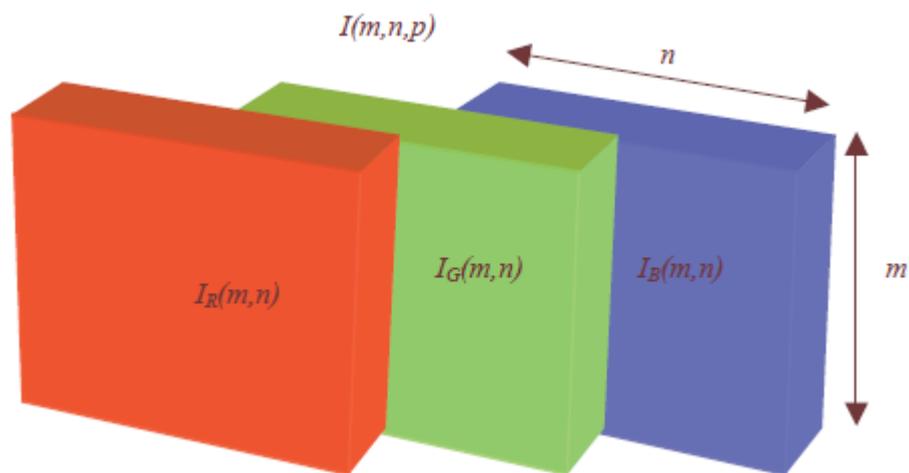


Figura 4: Representación de una imagen RGB.
Elaboración: Propia.

2.2.5.2. Imagen en escala de grises

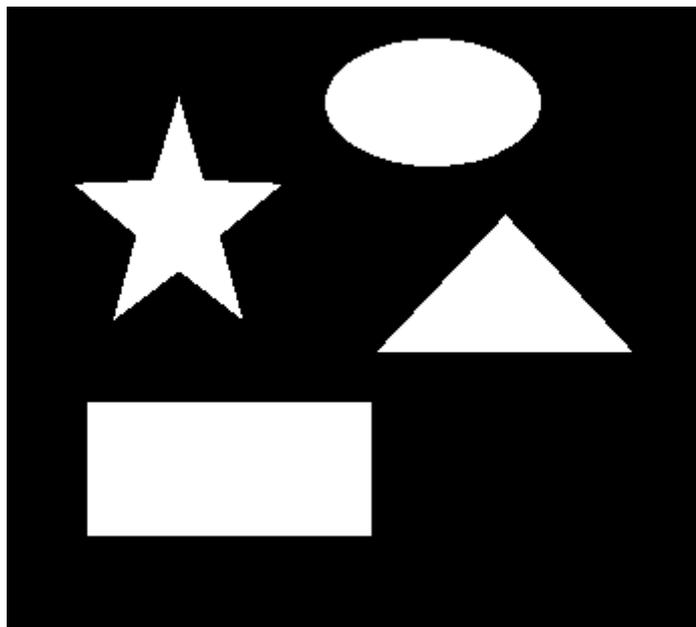
En la escala de grises una imagen digital llega a tener una tonalidad de 256 niveles que van desde 0 (negro) al 255 (blanco), el valor que tiene cada píxel es equivalente a una graduación de gris (ver figura 5).



**Figura 5: Representación de una en escala de grises.
Elaboración: Propia.**

2.2.5.3. Imagen binaria

Una imagen binaria es una imagen digital que tiene únicamente dos valores posibles 0 y 1 para cada píxel. Normalmente, los colores utilizados para su representación son negro y blanco (ver figura 6).



**Figura 6: Representación de una imagen binaria.
Elaboración: Propia.**

2.2.6. Clasificación de las imágenes

Por la forma de manejar los datos en un archivo de imagen, se puede hablar de dos modos principales para manipular la información que integra una imagen digital. Estos modos son las imágenes de mapa de bits y las imágenes vectoriales. Dado que cada uno se adapta mejor a un tipo de imagen; antes de conocer los diferentes formatos, debe conocerse el funcionamiento tanto de imágenes vectoriales como de imágenes de mapa de bits.

2.2.6.1. Imágenes vectoriales

Las imágenes vectoriales son gráficos formados a base de curvas y líneas a través de elementos geométricos independientes (segmentos, polígonos, etc) definidos por sus atributos matemáticos de forma por mencionar un ejemplo: un círculo de color verde quedaría definido por la posición de su centro, su radio, el grosor de línea y su color. La gran ventaja de las imágenes vectoriales es que no sufren pérdida de resolución al producirse una ampliación de los mismos (ver figura 7). Se utiliza mucho para trabajos de rotulación, rótulos, iconos, dibujos, logotipos de empresa etc. Esta clase de imagen tiene poco peso como archivo informático, medido en Kilobytes y tiene la particularidad de transformarse a mapa de bits si no se abren con el programa que los creó (Digitalfotored).

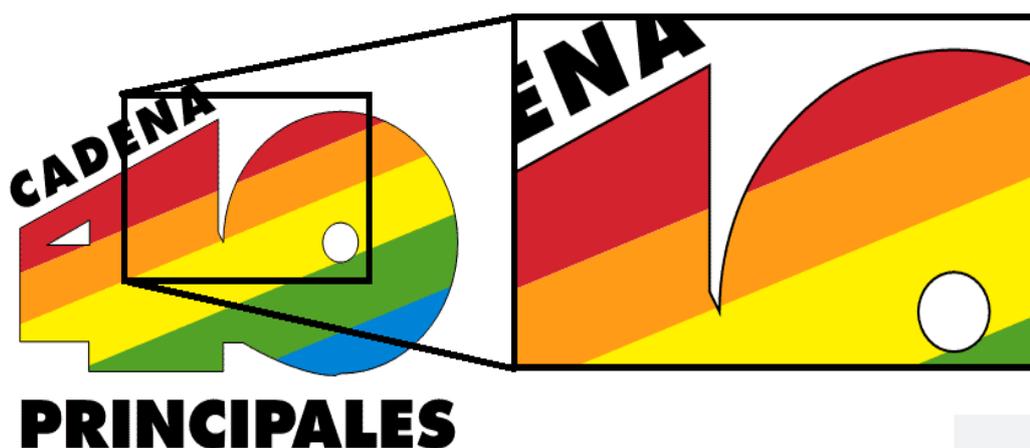


Figura 7: Imagen ampliada en un 200 % respecto al de su tamaño original.
Elaboración: (Digitalfotored).

2.2.6.2. Imágenes mapa de bits

Las imágenes de mapas de bits, denominadas técnicamente imágenes rasterizadas, utilizan una cuadrícula rectangular de elementos de imagen (píxeles) para representar imágenes. A cada píxel se le asigna una ubicación y un valor de color específicos. Al trabajar con imágenes de mapa de bits, se editan los píxeles, en lugar de los objetos o las formas. Las imágenes de mapa de bits son el medio electrónico más usado para las imágenes de tono continuo, como fotografías o pinturas digitales, puesto que pueden representar de manera más eficaz degradados sutiles de sombras y color.

Las imágenes de mapa de bits dependen de la resolución, es decir, contienen un número fijo de píxeles. Como consecuencia, pueden perder detalle y aparecer dentadas si se cambia la escala a grandes ampliaciones (ver figura 8) en pantalla o si se imprimen con una resolución inferior que aquella para la que fueron creadas (Adobe).



Figura 8: Imagen ampliada a diferentes niveles respecto al de su tamaño original. Elaboración: (Adobe).

2.2.7. Procesamiento digital de imágenes

El procesamiento digital de imágenes es el conjunto de técnicas que se aplican a las imágenes digitales con el objetivo de mejorar la calidad o facilitar la búsqueda de información.

Uno de los campos del procesamiento de imágenes digitales más aplicado y avanzado es la Visión Artificial o Visión por Computadora, la cual tiene como finalidad utilizar una computadora para emular la visión humana, incluyendo el poder de aprendizaje y la capacidad de realizar inferencias y acciones basadas en una entrada visual. Para ello, la visión por computadora hace uso de un software de reconocimiento de patrones, el cual permite clasificar objetos presentes en una imagen en diferentes clases establecidas convenientemente según el propósito (González & Woods, 2008). De esta manera, es posible discriminar diferentes formas de objetos y clasificarlos.

2.2.8. Detectores de borde

Un detector de borde permite una demarcación de los límites o bordes del objeto que se desea analizar; para ello, se detectan las discontinuidades significativas en los valores de intensidad de la imagen. Estas discontinuidades son detectadas usando la primera y segunda derivada. La primera utilizada en el procesamiento digital de imágenes es la gradiente (González & Woods, 2008).

La gradiente de una función de dos dimensiones, $f(x, y)$, es definido como el vector:

$$\nabla F = \begin{bmatrix} Gx \\ Gy \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (2.1)$$

La magnitud de este vector es:

$$\nabla f = \text{mag}(\nabla F) = [Gx^2 + Gy^2]^{1/2} \quad (2.2)$$

El vector gradiente apunta a la dirección del máximo cambio de f en las coordenadas (x, y) . El ángulo en la cual dicho máximo cambio ocurre es:

$$\alpha(x, y) = \tan^{-1} \left(\frac{Gy}{Gx} \right) \quad (2.3)$$

La segunda derivada en el procesamiento de imágenes es generalmente calculada usando el Laplaciano. La función $f(x, y)$ es la siguiente:

$$\nabla^2 f(x, y) = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} \quad (2.4)$$

El Laplaciano es raramente utilizado para la detección de borde porque, como segunda derivada, es muy sensible al ruido. su magnitud produce doble borde y no permite detectar la dirección del borde.

Las imágenes digitales manejan información discreta, para ello los detectores de borde utilizan máscaras para calcular el valor de la derivada de cada pixel analizando sus vecindades (ver figura 9).

Z_1	Z_2	Z_3
Z_4	Z_5	Z_6
Z_7	Z_8	Z_9

Figura 9: Representación de un pixel (Z_5) y sus ocho pixeles vecinos.
Elaboración: (González & Woods, 2008).

2.2.8.1. Detector de bordes de Roberts

El detector de bordes de Roberts se caracteriza por ser uno de los más simples y limitados, ya que no es simétrico y no permite detectar los bordes múltiples de 45° (González & Woods, 2008).

Utiliza las siguientes ecuaciones para aproximar las primeras derivadas G_x y G_y :

$$G_x = Z_9 - Z_5 \quad (2.5)$$

$$G_y = Z_8 - Z_6 \quad (2.6)$$

Con estos valores se determina la magnitud de la gradiente g del pixel:

$$g = [G_x^2 + G_y^2]^{1/2} \quad (2.7)$$

Para luego comparar el valor de la magnitud de la gradiente con un valor umbral T , de esta manera, si $g \geq T$, entonces es un pixel del borde del objeto.

2.2.8.2. Detector de bordes de Prewitt

El detector de borde de Prewitt utiliza las siguientes ecuaciones para aproximar digitalmente las primeras derivadas de derivadas G_x y G_y (Sato, 2013).

Utiliza las siguientes ecuaciones para aproximar las primeras derivadas G_x y G_y :

$$G_x = (Z_7 + Z_8 + Z_9) - (Z_1 + Z_2 + Z_3) \quad (2.8)$$

$$G_y = (Z_3 + Z_6 + Z_9) - (Z_1 + Z_4 + Z_7) \quad (2.9)$$

2.2.8.3. Detector de bordes de Sobel

El detector de borde de Sobel, para cada punto determina la magnitud del mayor cambio posible, utiliza coeficientes con valor 2, lo que le permite suavizar el ruido presente en la imagen:

Utiliza las siguientes ecuaciones para aproximar las primeras derivadas G_x y G_y :

$$G_x = (Z_3 + 2Z_6 + Z_9) - (Z_1 + 2Z_4 + Z_7) \quad (2.10)$$

$$G_y = (Z_7 + 2Z_8 + Z_9) - (Z_1 + 2Z_2 + Z_3) \quad (2.11)$$

2.2.8.4. Detector de bordes de Canny

El detector de borde de Canny es considerado uno de los más complejos y poderosos. Consiste en un algoritmo de múltiples etapas para detectar una amplia gama de bordes en imágenes; utiliza también la dirección de la gradiente, a diferencia de los otros ya mencionados que solo utilizan la magnitud. Este método emplea los siguientes pasos (González & Woods, 2008), (Tkalcic, 2003), (Sato, 2013):

1. La imagen es suavizada usando un filtro Gaussiano con una desviación estándar σ , para reducir el ruido.
2. Se calcula la magnitud de la gradiente y la dirección del borde para cada píxel.
3. Se realiza el proceso conocido como la supresión de no-máximos para obtener bordes más definidos y delgados. Para esto, se consideran cuatro direcciones identificadas por las orientaciones de 0° , 45° , 90° , 135° con respecto al eje horizontal y para cada píxel se encuentra la dirección que mejor se aproxima a la del gradiente. Si el valor de magnitud de la gradiente es más pequeño que al menos uno o de sus dos vecinos en la dirección del ángulo obtenido se asigna el valor de 0 a dicho píxel, en caso contrario se asigna el valor de la magnitud de la gradiente.
4. Se utiliza el concepto de histéresis con dos valores umbrales ($T1$ y $T2$) determinados para definir el borde. Si el valor de la magnitud de la gradiente es mayor a $T2$ es considerado un píxel de borde fuerte; si se encuentra entre los valores $T1$ y $T2$ es considerado un píxel de borde débil; y si es menor a $T1$ no es detectado como borde.
5. Finalmente se analiza la vecindad de los píxeles de borde débil para realizar la conexión con los píxeles de borde fuerte.

2.2.9. Operaciones morfológicas

Una operación morfológica es una herramienta que permite extraer componentes de una imagen que son útiles en la representación y descripción de la forma de una región como los bordes. Las operaciones morfológicas fundamentales son la dilatación y la erosión. Estas operaciones morfológicas y sus diferentes combinaciones son los que usualmente se utilizan en el procesamiento digital de imágenes. Las combinaciones más comunes de estas dos operaciones son la Apertura (Opening) y el Cierre (Closing) (González & Woods, 2008), (López, 2009).

2.2.9.1. Dilatación

La dilatación es una operación que hace crecer o engrosar los objetos en una imagen binaria. La forma específica y el alcance de este crecimiento son controlados por un elemento estructural. Si el origen o el centro del este último coincide con un pixel de valor 1 de la imagen, el elemento estructural solapa a la imagen original (González & Woods, 2008).

2.2.9.2. Erosión

La erosión es una operación morfológica que encoge o adelgaza los objetos en una imagen binaria. Del mismo modo que en la dilatación, la forma específica y el alcance de este encogimiento son controlados por un elemento estructural. Si todo el elemento estructural coincide con los valores 1 de la imagen, se obtiene en la salida solo la posición del origen del elemento estructural (González & Woods, 2008).

2.2.9.3. Apertura

También denominada Opening, esta operación morfológica se realiza aplicando una operación de erosión seguida de una dilatación. Sea la imagen **A** y el elemento estructural **B**, la operación de apertura se denota por $A \circ B$ y puede ser interpretada como la unión de todas las traslaciones de **B** que encajen completamente en **A**. A continuación se muestra el funcionamiento de la operación (ver figura 10) (González & Woods, 2008).



Figura 10: Ilustración de la operación morfológica apertura.
Elaboración: (González & Woods, 2008).

2.2.9.4. Cierre

La operación morfológica de cierre está dada por una operación de dilatación seguida de una erosión. Sea la imagen **A** y el elemento estructural **B**, la operación de cierre se denota por $A \bullet B$ y puede ser interpretada como el complemento de la unión de todas las traslaciones de **B** que no solapan a la imagen **A**. La siguiente muestra el funcionamiento de la operación (ver figura 11) (González & Woods, 2008).

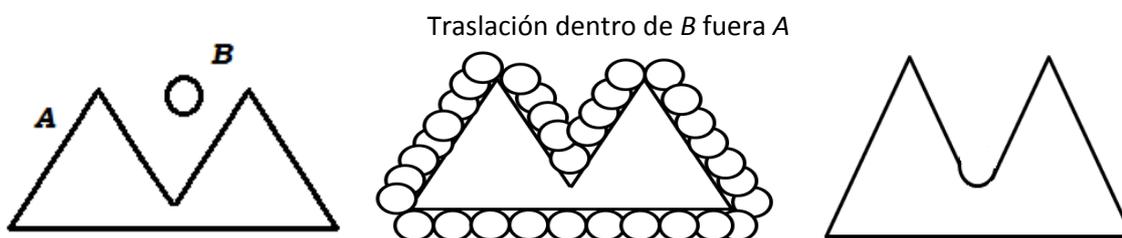


Figura 11: Ilustración de la operación morfológica cierre.
Elaboración: (González & Woods, 2008).

2.2.10. Descriptores de forma

Los descriptores de forma son herramientas comúnmente utilizadas en la etapa de extracción de características después de la etapa de segmentación de la imagen. Estos permiten obtener una representación numérica o matemática de las características de los objetos presentes en una imagen digital (Bow, 2002). Existen una gran variedad de tipos de descriptores entre los más utilizados son los Geométricos y los estadísticos, que pueden ser usadas de acuerdo al propósito del proyecto.

2.2.10.1. Descriptores geométricos

Los atributos geométricos son uno de los más simples y alguno de estos atributos no son invariantes a escala, traslación y rotación; sin embargo han demostrado brindar características suficientes para describir objetos y diferenciarlo correctamente a un costo computacional menor, a diferencia de los descriptores estadísticos. Algunos de los atributos más utilizados en el procesamiento digital de imágenes se muestran a continuación:

Perímetro: La aproximación más simple consiste en calcular el número de píxeles del borde de la imagen.

Es el número de píxeles que pertenecen al objeto y que, al menos, tienen un vecino que pertenece al fondo. La distancia entre píxeles en diagonal se suele tomar como $\sqrt{2} \approx 1,4142$.

Área: Es el número de píxeles que contiene una imagen.

Compacidad: Es una medida que no depende de las dimensiones de la región, por lo que es invariante por cambios de escala uniformes.

Bounding box: Es el rectángulo de menor tamaño que contiene al objeto.

2.2.10.2. Descriptores estadísticos

Los momentos: La geometría de una región plana se basa en el tamaño, la posición, la orientación y la forma. Todas estas medidas están relacionadas con una familia de parámetros llamada *momentos* (Shutler).

Los momentos de orden $p + q$ de una imagen $I(x, y)$ son:

$$M = \sum_{x=1}^N \sum_{y=1}^M x^p y^q I(x, y) \quad \text{Para } p \text{ y } q = 0, 1, 2, \dots \quad (2.12)$$

Siendo $I(x, y)$ el nivel de gris del píxel representado por el punto de coordenadas (x, y) . Por ejemplo, el momento de orden cero para una imagen binaria es el área del objeto:

$$M_{0,0} = \sum_{x=1}^N \sum_{y=1}^M I(x, y) \quad (2.13)$$

Los momentos centrales: Para poder hacer una descripción independiente de la posición del objeto, los momentos pueden calcularse respecto al centro de gravedad (Shutler).

Los *momentos centrales* son:

$$\mu_{p,q} = \sum_{x=1}^N \sum_{y=1}^M (x - \bar{x})^p (y - \bar{y})^q I(x, y) \quad (2.14)$$

Los momentos invariantes: Debido a que $\mu_{0,0}$ es el área del objeto, podemos normalizar los *momentos* para tener una descripción independiente del tamaño. Los *momentos centrales normalizados* son:

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^\alpha} \quad \alpha = \frac{p+q}{2} + 1 \quad (2.15)$$

A partir de los *momentos centrales normalizados* de orden dos y tres es posible extraer siete parámetros denominados momentos invariantes, independientes a posición, tamaño y ángulo del objeto (Theodoridis & Koutroumbas, 2009):

$$h_1 = \eta_{2,0} + \eta_{0,2} \quad (2.16)$$

$$h_2 = (\eta_{2,0} - \eta_{0,2})^2 + 4\eta_{1,1}^2 \quad (2.17)$$

$$h_3 = (\eta_{3,0} - 3\eta_{1,2})^2 + (3\eta_{1,2} - \eta_{3,0})^2 \quad (2.18)$$

$$h_4 = (\eta_{3,0} + \eta_{1,2})^2 + (\eta_{0,3} + \eta_{2,1})^2 \quad (2.19)$$

$$h_5 = (\eta_{3,0} - 3\eta_{1,2})(\eta_{3,0} + \eta_{1,2}) \left[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{0,3} + \eta_{2,1})^2 \right] + (3\eta_{2,1} - \eta_{0,3})(\eta_{0,3} + \eta_{2,1}) \left[3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{0,3} + \eta_{2,1})^2 \right] \quad (2.20)$$

$$h_6 = (\eta_{2,0} - \eta_{0,2}) \left[(\eta_{3,0} + \eta_{2,1})^2 - (\eta_{0,3} + \eta_{2,1})^2 \right] + 4\eta_{1,1}(\eta_{3,0} + \eta_{1,2})(\eta_{0,3} + \eta_{2,1}) \quad (2.21)$$

$$h_7 = (3\eta_{2,1} - \eta_{0,3})(\eta_{3,0} + \eta_{1,2}) \left[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{0,3} + \eta_{2,1})^2 \right] - (\eta_{3,0} - 3\eta_{1,2})(\eta_{0,3} + \eta_{2,1}) \left[3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{0,3} + \eta_{2,1})^2 \right] \quad (2.22)$$

2.2.11. Segmentación de imágenes

La segmentación es un proceso por el cual se divide una imagen digital en varias partes (grupos de píxeles) u objetos, con el fin de separar las partes de interés del resto de la imagen. Las imágenes segmentadas tienen características en

común como el color, brillo, forma, textura, movimiento, entre otros (La Serna & Román, 2009) (López, 2009). Estas técnicas están basadas en dos propiedades de los valores de intensidad:

- **Discontinuidad:** Se buscan cambios abruptos en la intensidad. Por ejemplo: extracción de bordes.
- **Similitud:** se buscan regiones similares según cierto criterio. Por ejemplo: umbralización del histograma, extracción de regiones, entre otros.

Existen varios tipos de segmentación, uno de ellos está basado en el método de la umbralización.

2.2.11.1 Umbralización

La umbralización es una técnica de simple y eficiente de segmentación que permite separar los píxeles de la imagen en escala de grises en dos categorías a partir de una ventana o borde umbral. El umbral puede ser función de la posición $p(r, c)$ de la vecindad $N(r, c)$ y de la intensidad $I(r, c)$ actual del píxel (Cortés, Chávez, & Mendoza, 2012).

$$T = T(p(r, c), N(r, c), I(r, c)) \quad (2.23)$$

Umbralización global

La umbralización global, es un valor único sobre toda la imagen (ver figura 12) sin tener en cuenta la vecindad. La imagen binaria $g(x, y)$ se obtiene a partir de la imagen original $f(x, y)$ en niveles de gris y de un umbral de separación (U) seleccionado mediante la siguiente condición (López, 2009):

$$g(x, y) = \begin{cases} 1 & \text{si } f(x, y) \geq U \\ 0 & \text{si } f(x, y) < U \end{cases} \quad (2.24)$$

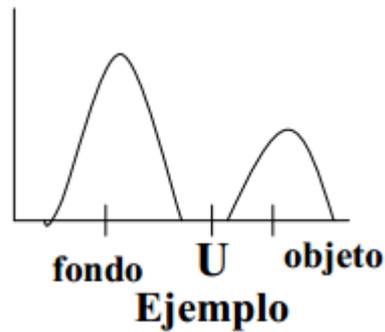


Figura 12: Ilustración del valor umbral (U) en el histograma de la imagen.
Elaboración: Propia.

La selección del umbral se basa en el histograma de la imagen donde se buscan y analizan los mínimos locales.

Este método resulta ser rápido, de coste computacional bajo. El método de Otsu es uno de los más importantes por las prestaciones que brinda.

Método de Otsu: El método de Otsu, llamado así en honor a Nobuyuki Otsu que lo inventó en 1979, utiliza técnicas estadísticas para resolver el problema. En concreto, se utiliza la variancia que es una medida de la dispersión de valores, en este caso se trata de la dispersión de los niveles de gris (IEEE, 1979).

El método de Otsu calcula el valor umbral de forma que la dispersión dentro de cada segmento sea lo más pequeña posible, pero al mismo tiempo la dispersión sea lo más alta posible entre segmentos diferentes. Para ello se calcula el cociente entre ambas varianzas y se busca un valor umbral para el que este cociente sea máximo (Cortés, Chávez, & Mendoza, 2012).

Una imagen es una función bidimensional de la intensidad del nivel de gris, y contiene N píxeles, cuyos niveles de gris se encuentran entre 1 y 255. El número de píxeles con nivel de gris i se denota como $f(i)$, y la probabilidad de ocurrencia del nivel de gris i en la imagen está dada por (Univeridad Nacional de Quilmes, 2005):

$$P_i = \frac{f(i)}{N} \tag{2.27}$$

En el caso de la umbralización en dos niveles de una imagen (a veces llamada binarización), los píxeles son divididos en dos clases: P_1 con niveles de gris $[1, \dots, t]$; y P_2 , con niveles de gris $[t + 1, \dots, 255]$.

En donde la distribución de probabilidades de cada clase se determina con la ecuación **(2.28)**:

$$P_1(t) = \sum_{i=0}^t p(i) \quad P_2(t) = \sum_{i=t+1}^{255} p(i) \quad (2.28)$$

Los promedios de clase se calculan con las ecuaciones en **(2.29)**:

$$\mu_1(t) = \frac{\sum_{i=0}^t i \cdot p(i)}{P_1(t)} \quad \mu_2(t) = \frac{\sum_{i=t+1}^{G-1} i \cdot p(i)}{P_2(t)} \quad (2.29)$$

Con las varianzas de cada clase presentadas mediante las ecuaciones en **(2.30)** y **(2.31)**:

$$\sigma_1^2(t) = \sum_{i=0}^t [i - \mu_1(t)]^2 \frac{p(i)}{P_1(t)} \quad (2.30)$$

$$\sigma_2^2(t) = \sum_{i=t+1}^{255} [i - \mu_2(t)]^2 \frac{p(i)}{P_2(t)} \quad (2.31)$$

La varianza entre clases a través de la ecuación **(2.32)**:

$$\sigma_w^2 = P_1(t)\sigma_1^2(t) + P_2(t)\sigma_2^2(t) \quad (2.32)$$

Para una umbralización de dos niveles, Otsu verificó que el umbral óptimo t^* se elige de manera que σ_w^2 sea máxima; esto es:

$$t^* = \text{Max}_t \{ \sigma_w^2(t) \} \quad 1 \leq t \leq 255 \quad (2.33)$$

Umbralización local

La umbralización local depende de las características locales de la imagen y las coordenadas espaciales x e y , es por eso que no hay un único valor umbral para la imagen, pueden haber múltiples umbrales (La Serna & Román, 2009).

$$g(x, y) = \begin{cases} 1 & \text{si } f(x, y) \geq T(x, y) \\ 0 & \text{si } f(x, y) < T(x, y) \end{cases} \quad (2.34)$$

Dónde:

$$T(x, y) = f_0(x, y) + T_0 \quad (2.35)$$

Este método es recomendable para corregir problemas de iluminación en la imagen.

2.2.12. Reconocimiento de patrones

Consiste en el reconocimiento de patrones de señales. Los patrones se obtienen a partir de los procesos de segmentación, extracción de características y descripción donde cada objeto queda representado por una colección de descriptores. El software de reconocimiento debe asignar a cada objeto su categoría o clase (conjunto de entidades que comparten alguna característica que las diferencia del resto). Para poder reconocer los patrones se siguen los siguientes procesos (Meyer):

- Adquisición de datos
- Extracción de características
- Toma de decisiones

En forma esquemática el proceso de reconocimiento de patrones se muestra en la figura 13.

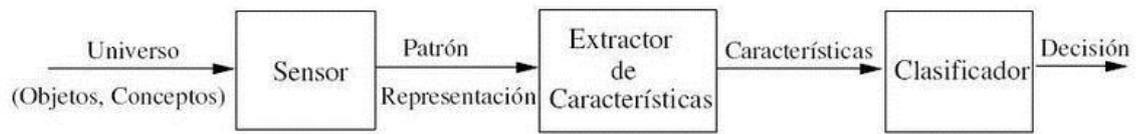


Figura 13: Proceso de reconocimiento de patrones.

Elaboración: (Meyer).

“Un patrón es una descripción estructural o cuantitativa de un objeto o de alguna otra entidad de interés en una imagen.”

“En general, un patrón está formado por uno o más descriptores. Las clases de patrones se representan por W_1, W_2, \dots, W_m , donde m es el número de clases. Las tres representaciones de patrones más utilizadas son: los vectores (para descripciones cuantitativas), las cadenas y los árboles (para descripciones estructurales)” (Romo, Ramírez, & Valdivieso, 2007).

La esencia del reconocimiento de patrones es la discriminación de una señal, dependiendo de sus características.

CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN

3.1. TIPO DE INVESTIGACIÓN

La presente investigación tiene por objetivo desarrollar un software que optimice el tiempo y reduzca el margen de error en la discriminación de granos de arroz pulido. Es la aplicación de los conocimientos adquiridos para desarrollar esta herramienta. Se enfoca en la solución de problemas en una realidad concreta. Por lo tanto es una **Investigación Aplicada**.

3.2. MÉTODO DE INVESTIGACIÓN

El Método Empírico. Se caracteriza por adquirir conocimiento en base a la experiencia. El procedimiento es el siguiente: hecho, la observación, la medición y el experimento. Teniendo conocimiento de la realidad del problema y la solución planteada, se realizan los experimentos con las condiciones mostradas en el presente trabajo, para luego comparar los resultados obtenidos por el software y el método convencional.

3.3. ALCANCE DE LA INVESTIGACIÓN

La presente investigación tiene un **Estudio de Alcance Explicativo**, en donde se crea las condiciones necesarias para desarrollar un experimento, en donde se manipula la variable independiente de manera deliberada en este caso es el “Software automático basado en procesamiento digital de imágenes”, para observar cual es el efecto que tiene sobre la variable dependiente (ver punto 3.7. Sistema de Variables).

3.4. DISEÑO DE INVESTIGACIÓN

El diseño de la investigación hace referencia a la estrategia concebida para obtener información. La presente investigación tiene un **diseño experimental** por la manipulación intencional que se realiza a la variable independiente que alcanza dos niveles: presencia y ausencia.

Para realizar una adecuada contrastación de resultados, se lleva a cabo varios tipos de experimentos, que se hacen con muestras elegidas de manera aleatoria para el tipo “experimento puro” y muestras con características conocidas a priori para el tipo “cuasi experimento”.

Cabe señalar que al hablar de grano o granos en los experimentos, se hace referencia a granos de arroz pulido comerciales.

Experimento Nro 01.

El experimento se realiza con 01 (una) muestra de 10g de arroz pulido elegidos al azar. Se realizará una observación con el método manual, y luego otra observación con el estímulo, para posteriormente comparar los resultados obtenidos.

El diagrama es como se muestra a continuación:

RG_1 $-$ O_1 X O_2

Experimento Nro 02.

El experimento se realiza con 01 (una) muestra de 215 granos de arroz pulido, en la que la información de las características de clase y tipo de defecto se conoce de antemano. Se ha realizado 03 observaciones por cada clase para verificar cuántos son los granos, que por lo menos, aparecen en dos clases o tipo de defecto diferentes. Esto con el objetivo de someter al sistema a prueba para determinar su confiabilidad en la discriminación.

El diagrama es de la siguiente manera:

$$G_1 \quad X \quad O_1 \quad O_2 \quad O_3$$

Experimento Nro 03.

El experimento se realiza con 01 (un) grano elegido al azar y medirlo 10 (diez) veces, con el objetivo de determinar el margen de error en la medición.

$$RG_1 \quad X \quad O_1 \quad O_2 \quad O_3 \quad O_4 \quad O_5 \quad O_6 \quad O_7 \quad O_8 \quad O_9 \quad O_{10}$$

$$RG_2 \quad - \quad O_{11} \quad O_{12} \quad O_{13} \quad O_{14} \quad O_{15} \quad O_{16} \quad O_{17} \quad O_{18} \quad O_{19} \quad O_{20}$$

3.5. POBLACIÓN Y MUESTRA DE ESTUDIO

3.5.1. Población

El universo poblacional está conformado por un lote de granos de arroz pulido comercial de calidad extra con una masa de 45 Kg.

3.5.2. Muestra

Las muestras utilizadas fueron probabilísticas, en donde todos los elementos tuvieron la misma probabilidad de ser elegidos. Fueron escogidos por números aleatorios generados por una computadora y sorteos.

Cada muestra está conformada por una masa de 10g que es una cantidad elegida de manera arbitraria.

El presente trabajo no tiene como objetivo realizar una inferencia estadística sobre el lote de granos.

3.6. SISTEMA DE HIPÓTESIS

3.6.1. Hipótesis principal

La aplicación del software automático basado en procesamiento digital de imágenes reduce el tiempo en el conteo total de granos de arroz pulido y en el

proceso de discriminación por su clase y tipo de defecto, así como disminuye el margen de error en la medición de longitud.

3.6.2. Hipótesis secundarias

Hipótesis secundaria 1

El Software automático basado en procesamiento digital de imágenes reduce el tiempo en el conteo de los granos de arroz pulido.

Hipótesis secundaria 2

El Software automático basado en procesamiento digital de imágenes reduce el tiempo en la discriminación de los granos de arroz pulido según su clase.

Hipótesis secundaria 3

El Software automático basado en procesamiento digital de imágenes reduce el tiempo en la discriminación de los granos de arroz pulido según su tipo de defecto.

Hipótesis secundaria 4

El Software automático basado en procesamiento digital de imágenes reduce el margen de error en la medición de longitud de los granos de arroz pulido.

3.7. SISTEMA DE VARIABLES

3.7.1. Variable independiente

X: Software automático basado en procesamiento digital de imágenes.

3.7.2. Variable dependiente

Y1: Tiempo en el conteo de los granos de arroz pulido.

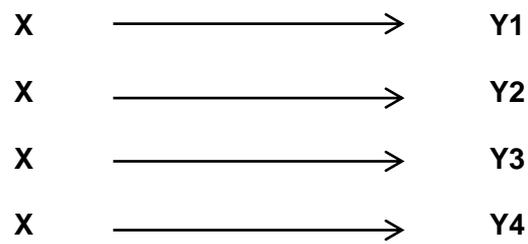
Y2: Tiempo en la discriminación de los granos de arroz pulido según su clase.

Y3: Tiempo en la discriminación de los granos de arroz pulido según su tipo de defecto.

Y4: Margen de error en la medición de los granos de arroz pulido.

3.7.3. Relación entre las variables

A continuación se muestra la relación que tienen las variables:



CAPÍTULO IV: DESARROLLO DE LA SOLUCIÓN

4.1. PLAN DE DESARROLLO DEL SOFTWARE

La planificación del software desarrollado consistió en realizar una evaluación de la tecnología, el costo económico y recursos humanos que implica la implementación del software, así como la relevancia y el impacto que tendrá su funcionamiento. Del mismo modo, estuvo destinado a recolectar información sobre los componentes que serán utilizados en el desarrollo del software.

4.1.1. Beneficio

Las personas beneficiadas por las bondades que ofrece el software, son todas aquellas que laboran con granos de arroz pulido y necesitan saber las características morfológicas para su discriminación en función a su clase y tipo de defecto de cada uno de ellos y su distribución porcentual en una muestra, con un bajo coste computacional. Los resultados se obtienen de manera automática, optimizando el tiempo empleado en la discriminación y reduciendo el margen de error.

4.1.1.1. Beneficios tangibles

Los beneficios tangibles aportados por el sistema propuesto están dados en los siguientes aspectos:

- Reducción de costos en pagos y capacitación al personal.
- Reducción cantidad horas/hombre empleadas para este proceso.
- Reducción del espacio físico empleado para este proceso.
- Ahorro en suministros para los equipos del personal.

4.1.1.2. Beneficios intangibles

Entre los beneficios intangibles del sistema propuesto se pueden incluir:

- Optimizar el proceso de discriminación de granos de arroz pulido, aumentando la productividad del personal que labora en el mismo.
- Crear un adecuado control y seguimiento de los registros (imágenes digitales procesadas) de los granos de arroz pulido.
- Ofrecer la capacidad de discriminar gran volumen y diversidad de granos de arroz pulidos con rapidez y precisión. Ofrece una herramienta tecnológica de trabajo al personal, para facilitar sus labores.
- Generar información más precisa y confiable, que sirva de apoyo a la toma de decisiones.
- Mayor y mejor aprovechamiento de los recursos tecnológicos.
- Capacidad de registrar y almacenar “automáticamente” datos de los registros, estandarizando el mantenimiento de los registros, lo que implica un aumento de la capacidad y seguridad de almacenamiento de registros.

4.1.2. Hardware y software empleado

El hardware y software empleado para el desarrollo y funcionamiento del sistema son los siguientes:

Hardware:

- 01 Laptop Samsung 300E4C:
 - Core i3 2.4Ghz (4 CPUs)
 - Memoria RAM 4.00 GB
 - Software operativo Windows 7 Ultimate 64 bits
 - Disco Duro 750 GB
 - Tarjeta de video de 1.5 GB
- 01 Scanner CanoScan LiDE 110
 - Puerto de conexión USB
 - Fuente de luz : LED de 3 colores

- Resolución óptica: 2400 x 4800 dpi/ppp
- Resolución seleccionable: 19.200 dpi/ppp
- Profundidad de bits de escaneado: entrada de 48 bits(16 bits para cada color RGB)
- Tamaño máximo de escaneo desde el cristal: 21,6 x 29,7 cm (8,5 x 11,7 pulgadas).

Software:

- MatLab R2010a y su Toolbox Image Processing.

4.2. ESTUDIO DE FACTIBILIDAD

Teniendo definida la problemática se procedió a establecer las causas que ameritan la implementación de software. Es pertinente realizar un estudio de viabilidad para determinar la infraestructura tecnológica y la capacidad técnica que implica el desarrollo del software propuesto, así como los costos, beneficios y el grado de aceptación que la propuesta genera en los usuarios.

Este análisis permite determinar las posibilidades de diseñar el software propuesto. Los aspectos tomados en cuenta para este estudio fueron clasificados en tres áreas, las cuales se describen a continuación:

4.2.1. Viabilidad técnica

La factibilidad técnica consistió en realizar una evaluación de la tecnología existente en el mercado. Este estudio estuvo destinado a recolectar información sobre los componentes técnicos que serán utilizados en el desarrollo del software propuesto, de ser necesario los requerimientos tecnológicos deberán ser adquiridos.

De acuerdo a la tecnología necesaria para el desarrollo e implementación del software automático de procesamiento digital de imágenes para la discriminación

de granos de arroz pulido basado en procesamiento digital de imágenes, se evaluó bajo dos enfoques:

4.2.1.1. Hardware

En cuanto al hardware, específicamente el computador donde se desarrolló e implementó el software, este tiene los requerimientos mínimos para su funcionamiento. Además, se tiene disponible un escáner para la obtención de imágenes y un vernier para medir la longitud de los granos, implementos necesarios para el desarrollo de la presente investigación.

4.2.1.2. Software

En cuanto al software, se cuenta con todas las aplicaciones necesarias para el desarrollo del proyecto y funcionamiento del mismo, no amerita inversión alguna para su adquisición. El computador cuenta con el software mínimo para su funcionamiento, posee las herramientas de escritorio que existen en el mercado actualmente. El desarrollo de la aplicación y procesamiento digital de imágenes se realizará con el programa MatLab R2010a.

Como resultado de este estudio técnico se determinó que a la fecha se cuenta con la infraestructura tecnológica (Hardware y Software) necesaria para la implementación del software.

4.2.2. Viabilidad económica

A continuación se presenta un estudio que dio como resultado la factibilidad económica del desarrollo del software. Se determinaron los recursos para desarrollar el software en mención, haciendo una evaluación donde se puso en manifiesto el equilibrio existente entre los costos del sistema y los beneficios que se derivó de este, lo cual permitió observar de una manera más precisa las bondades del sistema propuesto.

A continuación se presenta un resumen de los costos que involucra el desarrollo e implantación del sistema propuesto:

4.2.2.1. Costos Generales

Al lograr optimizar el proceso de discriminación de granos de arroz pilado, se redujo el tiempo y error del mismo. También se disminuye la cantidad de personal que normalmente sería necesario para cumplir esta labor.

4.2.2.2. Costos de Hardware y Software

Debido a que actualmente se cuenta con los equipos y recursos tecnológicos necesarios, para el desarrollo e implementación del software, no fue requerido ningún tipo de inversión en este aspecto.

El software en el que se realizó la programación fue MatLab R2010a con una licencia Test drive Image Processing Toolbox prueba gratuita de 30 días.

4.2.2.3. Costos de personal

El equipo de desarrollo del software conformado por el asesor y el investigador, no generó inversión considerable, por tratarse de un proyecto de investigación para optar el título profesional. No existe un lucro de por medio.

4.2.3. Viabilidad operativa

La necesidad de los productores de granos de arroz es evidente porque es de importancia para ellos optimizar este proceso, ya que demanda tiempo y costo. La aceptación del sistema que automatiza la discriminación de granos de arroz pulido por parte de los productores y comerciantes es positiva, ya que su proceso de discriminación es manual y poco automatizado. Por otro lado, el software se usará de una manera más sencilla y amigable, que cubrirá todos sus requerimientos, expectativas y proporciona la información en forma oportuna y confiable. Basándose en la Norma Técnica Peruana (INDECOPI, 2012), Establecimiento de Identidad y de Calidad de Arroz Elaborado (MERCOSUR), conversaciones sostenidas con el personal involucrado se demostró que estos no representan ninguna oposición al cambio, por lo que el sistema es factible operacionalmente.

Con la finalidad de garantizar el buen funcionamiento del sistema y que este impactará en forma positiva a los usuarios, el mismo fue desarrollado en base a las normas técnicas que regulan la producción y comercialización de granos de arroz, presentando una interfaz amigable al usuario, lo que se traduce en una herramienta de fácil manejo y comprensión; tanto las pantallas como los reportes serán familiares a los operadores, contando con la opinión de los mismos para cualquier modificación del sistema.

4.2.4. Relación costo beneficio

El Análisis Costo-Beneficio presenta grandes ventajas para los usuarios, ya que se necesitan requisitos mínimos (hardware y software) para el desarrollo e implantación del software, por lo que no se necesitó grandes sumas de dinero en lo que a tecnología se refiere.

El sistema propuesto genera mejoras significativas en la reducción de tiempo y error en la discriminación de granos de arroz pulido. Es una herramienta efectiva que disminuyó las cargas de trabajo a los operadores, ya que la velocidad de procesamiento, veracidad y confiabilidad de los procesos y resultados serán los deseados.

Se toma en cuenta el valor que tiene la información, siendo el punto de apoyo en el proceso de la toma de decisiones. Las organizaciones deben su éxito en gran parte a la importancia que le dieron a la información y que estas han otorgado el verdadero valor que debe tener dentro de sus procesos.

4.3. DESARROLLO

4.3.1. Metodología de desarrollo

La metodología empleada en el desarrollo del software fue OpenUP, por ser una metodología ágil y un proceso mínimo y suficiente, lo que significa que solo el contenido fundamental y necesario es incluido. Las fases que lo componen se muestran a continuación:

4.3.1.1. Inicio

En esta primera fase de la metodología se realiza el primer contacto con la problemática, en donde se abstrae la realidad, para determinar cuál es la tecnología que puede solucionar los problemas. Se identifican cuáles son los procesos que deben automatizarse y de qué manera se logrará esto.

Una parte importante es definir claramente los alcances y las limitaciones que tendrá el software, los que ya fueron explicados con anterioridad.

4.3.1.2. Elaboración

Esta es la fase en donde se tiene un buen entendimiento acerca de los requisitos que debe cumplir el software, en este caso vendría a ser determinar el número total de granos de arroz pulido en una imagen digital y discriminarlos en función a su clase y tipo de defecto. También se diseña la interfaz a desarrollar en donde son implementadas y probadas aunque carecen de las funcionalidades.

4.3.1.3. Construcción

Esta es la etapa que se orienta totalmente en la construcción de la aplicación o programación. Como se mencionó anteriormente el programa que utilizaremos para la programación es MatLab R2010a con su herramienta Toolbox Image Processing. Inicialmente se construyó una versión beta, pero a medida que el ciclo fué iterando esta versión maduró. Más adelante se dará una explicación detallada acerca de la programación.

4.3.1.4. Transición

El desempeño y la calidad son puestos a prueba en esta fase, asegurándonos de que el software está listo para entregarse a los usuarios. Esta fase es muy importante donde se depuran potenciales errores, y donde se someten a distintos experimentos que se explicaron en el apartado anterior (3.4), para determinar su nivel de confiabilidad y resultados que se explica en el punto siguiente (5.2) y satisfacer las necesidades del usuario.

4.3.2. Análisis de requerimientos

Es el proceso por el cual se establece los servicios que el consumidor requiere de un sistema y las restricciones sobre las cuales debe funcionar y ser desarrollado.

4.3.2.1. Requerimientos funcionales

A continuación se describe las funcionalidades y servicios del software.

- El software deberá poder abrir imágenes digitales en formato “*.jpg” y mostrarlo en su interfaz gráfica (RF-01).
- El software a través del procesamiento digital de imágenes deberá poder binarizar la imagen digital (RF-02).
- El software a deberá poder determinar la cantidad de granos de arroz que tiene una imagen digital.
- El software a deberá poder discriminar los granos de arroz de una imagen digital en función a su clase (grano grande, mediano o corto) y mostrar otra imagen con el perímetro del grano pintado (verde, amarillo o rojo respectivamente) (RF-04).
- El software a través del procesamiento digital de imágenes deberá poder reducir el ruido del fondo de la imagen digital (RF-05).
- El software a deberá poder discriminar los granos de arroz de una imagen digital en función a su tipo de defecto (grano normal, tizoso total o tizoso parcial) y mostrar otra imagen con el perímetro del grano pintado (verde, rojo o amarillo respectivamente) (RF-03).
- El software a deberá poder grabar la imagen modificada en un formato “*.jpg” (RF-06).

4.3.2.2. Requerimientos No Funcionales

A continuación se definen las propiedades y restricciones del software a construir.

- La interfaz debe ser implementada en MatLab R2010a.
- El tiempo de actualización de la pantalla inferior a 2 segundos al cargar la imagen o mostrar su procesamiento digital.
- La aplicación deberá funcionar sobre un sistema operativo Windows.
- Las imágenes deben ser escaneadas a 300ppp.

4.3.3. Especificaciones de los casos de uso

A continuación se muestran la descripción del actor y los casos de uso:

4.3.3.1. Descripción de actores

ACT-01	Operario
Descripción	Este actor representa al usuario del sistema
Comentarios	Ninguno

Tabla 3: Especificación del actor Operario.
Elaboración: Propia.

4.3.3.2. Descripción de casos de uso

Software automático de procesamiento digital de imágenes	
DESCRIPCION DE CASOS DE USO	
Nombre:	LeerImagen
Alias:	
Actores:	Operario
Función:	Permitir la apertura de una imagen.
Descripción:	El responsable puede abrir imágenes, para su procesamiento digital y su discriminación.
Referencia:	De requerimiento: RF-01

Tabla 4: Descripción del caso de uso LeerImagen.
Elaboración: Propia.

Software automático de procesamiento digital de imágenes	
DESCRIPCION DE CASOS DE USO	
Nombre:	BinarizarImagen
Alias:	
Actores:	Operario
Función:	Efectuar la binarización de una imagen.
Descripción:	El responsable puede binarizar las imágenes convirtiéndolas en blanco y negro.
Referencia:	De requerimiento: RF-02

Tabla 5: Descripción del caso de uso BinarizarImagen.
Elaboración: Propia.

Software automático de procesamiento digital de imágenes	
DESCRIPCION DE CASOS DE USO	
Nombre:	DiscriminarTipo
Alias:	
Actores:	Operario
Función:	Efectuar la discriminación por tipo.
Descripción:	El responsable puede realizar la discriminación de granos de arroz pulido de una imagen digital en función a su tipo de defecto (grano normal, tizoso total y tizoso parcial).
Referencia:	De requerimiento: RF-03

Tabla 6: Descripción del caso de uso DiscriminarTipo.
Elaboración: Propia.

Software automático de procesamiento digital de imágenes	
DESCRIPCION DE CASOS DE USO	
Nombre:	DiscriminarClase

Alias:	
Actores:	Operario
Función:	Efectuar la discriminación por clase.
Descripción:	El responsable puede realizar la discriminación de granos de arroz pulido de una imagen en función a su clase (grano largo, mediano y corto).
Referencia:	De requerimiento: RF-04

Tabla 7: Descripción del caso de uso DiscriminarClase.

Elaboración: Propia.

Software automático de procesamiento digital de imágenes	
DESCRIPCION DE CASOS DE USO	
Nombre:	ReducirRuido
Alias:	
Actores:	Operario
Función:	Efectuar el procesamiento digital.
Descripción:	El responsable puede realizar operaciones sobre la imagen y normalizar el fondo.
Referencia:	De requerimiento: RF-05

Tabla 8: Descripción del caso de uso ReducirRuido.

Elaboración: Propia.

Software automático de procesamiento digital de imágenes	
DESCRIPCION DE CASOS DE USO	
Nombre:	GrabarImagen
Alias:	
Actores:	Operario

Función:	Permite grabar la imagen.
Descripción:	El responsable puede grabar la imagen modificada.
Referencia:	De requerimiento: RF-05

Tabla 9: Descripción del caso de uso GrabarImagen.

Elaboración: Propia.

4.3.4. Diagrama de los casos de uso

A continuación se muestra el diagrama de casos de usos del software desarrollado:

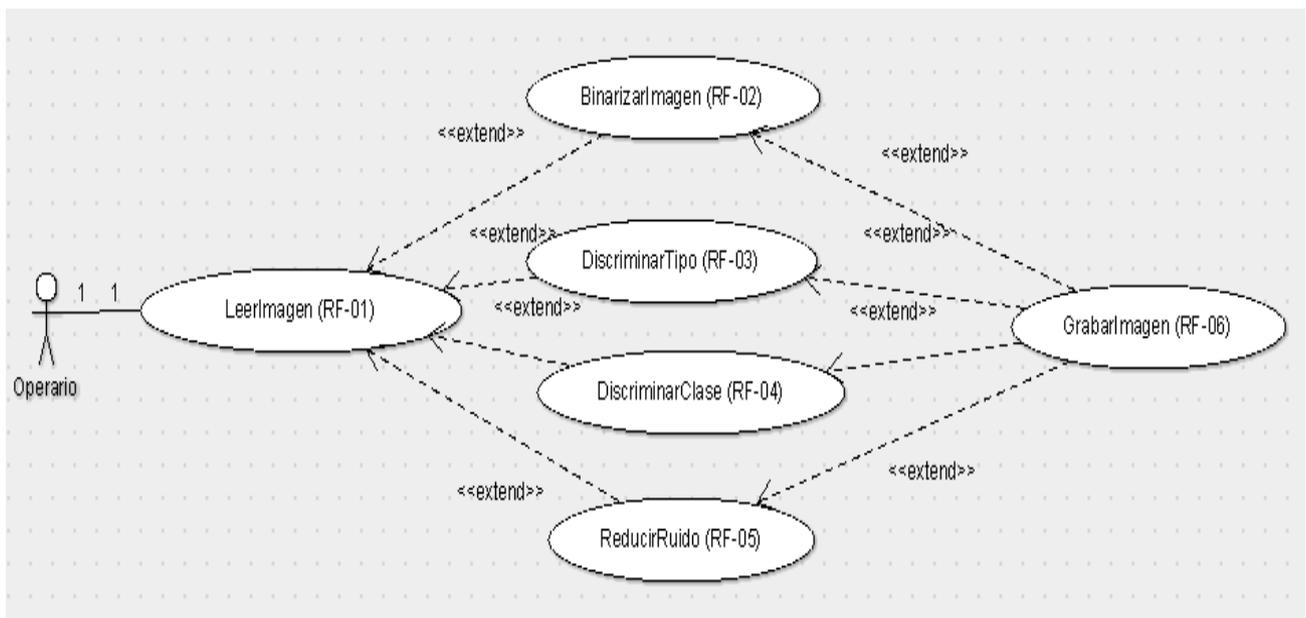


Figura 14: Diagrama de los casos de uso.

Elaboración: Propia.

4.4. DESCRIPCIÓN DE LA SOLUCIÓN DESARROLLADA

El software automático de reconocimiento de patrones para la discriminación de granos de arroz pulido basado en procesamiento digital de imágenes fue desarrollado bajo la metodología de desarrollo de software OpenUp, y haciendo uso del programa MatLab R2010a con el Toolbox Image Processing.

Para el desarrollo del algoritmo de procesamiento digital de imágenes se planteó el siguiente esquema, tal como se muestra en la siguiente figura:

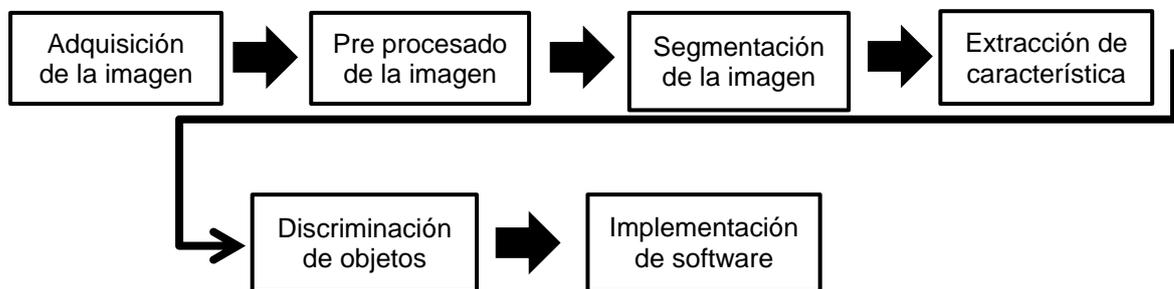


Figura 15: Diagrama de bloques de la solución planteada.

Elaboración: Propia.

A continuación se detallan cada uno de estos pasos:

4.4.1 Adquisición de la imagen

Es esta primera etapa se coloca los granos de arroz pulido en el escáner para obtener una imagen digital en un formato “*.jpg” con una composición de colores RGB a 300ppp (ver figura 16). En la imagen escaneada se observa los siguientes problemas:

- Los granos al encontrarse en un fondo de color blanco dificultan la determinación de sus características por tener un color semejante y requieren un mayor tratamiento digital, esto se refleja en un mayor coste computacional.
- El escáner ilumina el objetivo a través del cristal, y como todo objeto genera una sombra en el fondo, se muestra en la figura 16, después del procesamiento digital la sombra aparenta ser parte física del grano,

por la variación de color (de la sombra y el fondo), esto causa una distorsión en el resultado.

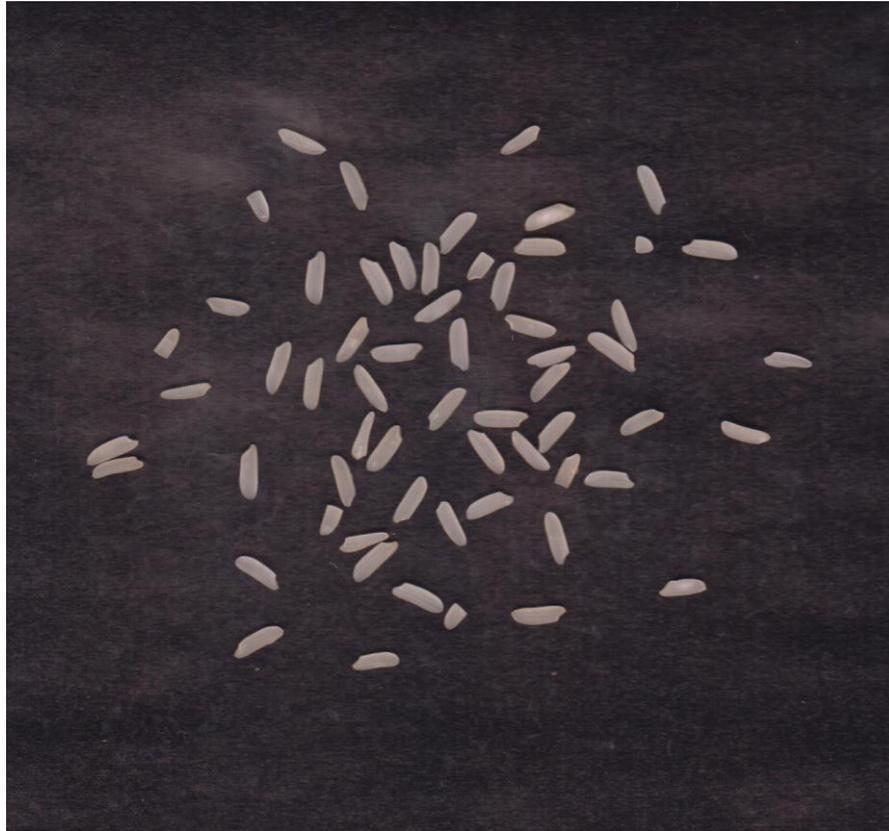


**Figura 16: Imagen escaneada de granos de arroz pulido con fondo blanco.
Elaboración: Propia.**

Nuestro conocimiento empírico sugiere que de ser posible es mejor tener un mayor contraste entre el fondo y los objetos de estudio en la imagen (ver figura 17).

Es por ese motivo que se adecuó una hoja de color negro en el fondo del escáner. Para eliminar la sombra del grano e incrementar el contraste. Así se reduce si significativamente el coste computacional.

Se requiere una imagen digital por cada una de las muestras para su posterior procesamiento digital y discriminación.



**Figura 17: Imagen escaneada de granos de arroz pulido con fondo negro.
Elaboración: Propia.**

4.4.2. Pre procesado de la imagen

Una vez obtenida la imagen digital se procede a realizar un pre procesamiento para eliminar el ruido y ciertas partes de la imagen que son irrelevantes como el fondo de la imagen, para ello se hará uso del software MatLab R2010a, debido a que dicho software cuenta con una alta prestación para el tratamiento de señales e imágenes digitales, además se sabe que es un programa de cálculo numérico orientado a matrices, y como se vio anteriormente, las imágenes digitales son un arreglo de matrices.

En este proceso se lleva la imagen digital con tonalidad RGB a una imagen binarizada compuesta por 0s y 1s (Negro y blanco respectivamente), para determinar el valor umbral se utiliza el método de Otsu que elige el umbral T .

De acuerdo a la siguiente función:

$$g(x, y) = \begin{cases} 1, & f(x, y) > T \\ 0, & f(x, y) \leq T \end{cases} \quad (4.1)$$

El umbral elegido debe ser tal que en la porción de la imagen que ha de ser analizada no se pierda muchos datos después de la umbralización; es decir, se debe mantener la mayor parte de los datos (área del grano) de dicha porción de la imagen; sin embargo, esto ocasionará que la imagen umbralizada obtenga porciones de ruido, los cuales luego deben ser eliminados (ver figura 18).

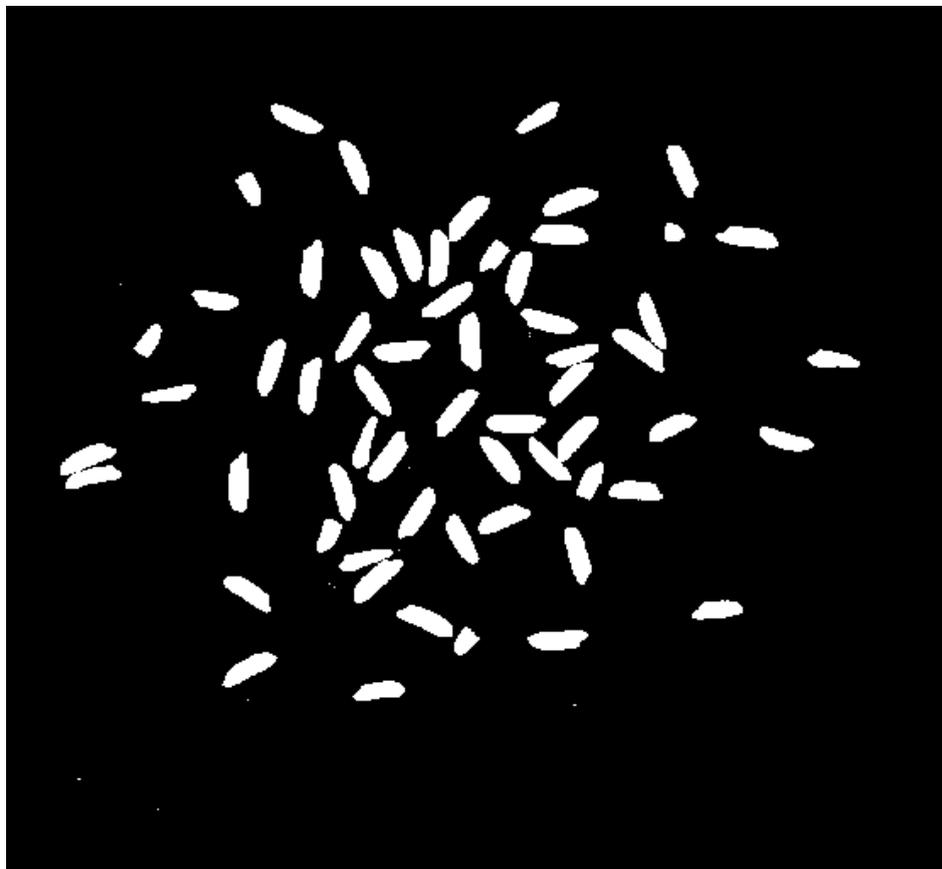
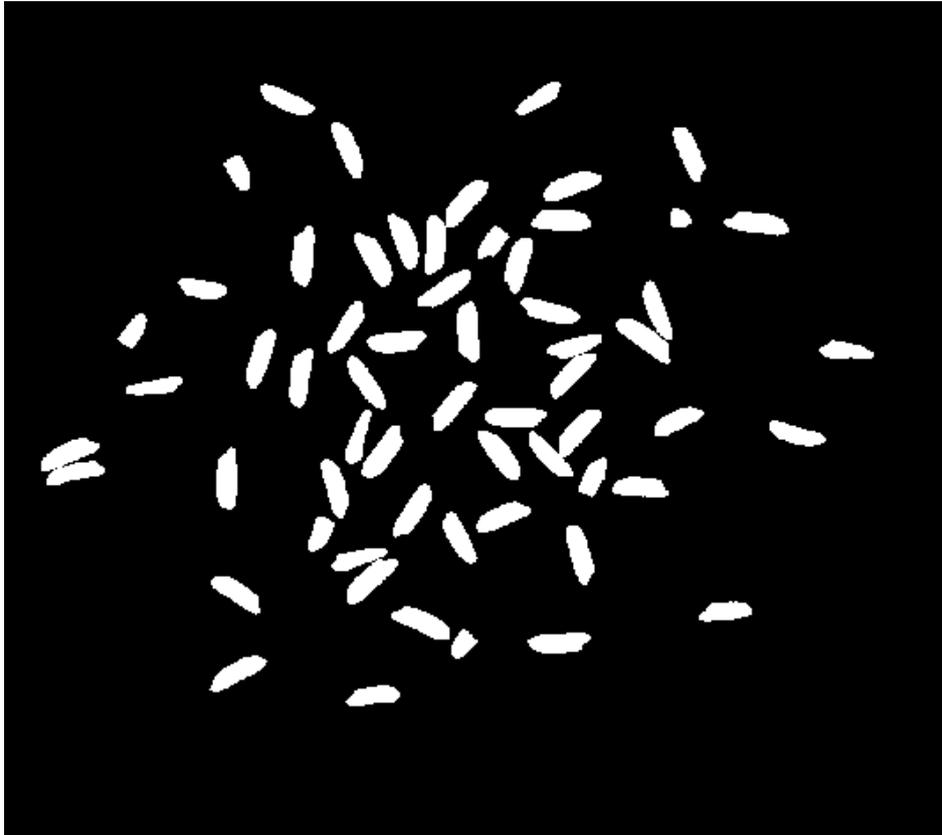


Figura 18: Imagen binarizada con umbral obtenido con el método de Otsu.
Elaboración: Propia.

Como ya se mencionó, se observa ruido en el fondo de la imagen (ver figura 18), áreas que no son granos; sin embargo, pintan como ellos. El software procede a eliminar las áreas pequeñas como se muestra a continuación (ver figura 19).



**Figura 19: Imagen binarizada donde se eliminó el ruido por completo.
Elaboración: Propia.**

La imagen está constituida por una matriz con 0s y 1s como elementos, a través de multiplicaciones matemáticas se filtra la imagen original y se muestra sin el ruido del fondo (ver figura 20 y 21) en una tonalidad RGB.

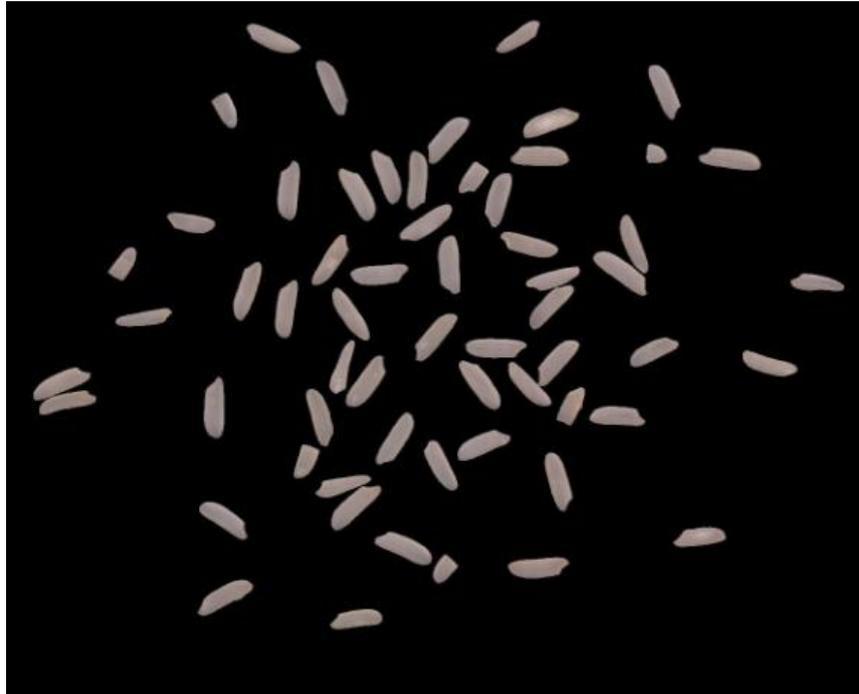


Figura 20: Imagen RGB donde se eliminó el ruido en el fondo (negro) por completo.

Elaboración: Propia.



Figura 21: Imagen RGB donde se eliminó el ruido en el fondo (blanco) por completo.

Elaboración: Propia.

4.4.3. Segmentación de la imagen

Una vez obtenida una imagen filtrada adecuadamente procedemos a etiquetar los elementos que se encuentran y segmentar cada uno de estos, para obtener solo la parte de dicha imagen que es necesaria para el análisis. Algunos de los granos segmentados se observan en la siguiente figura (ver figura 22):

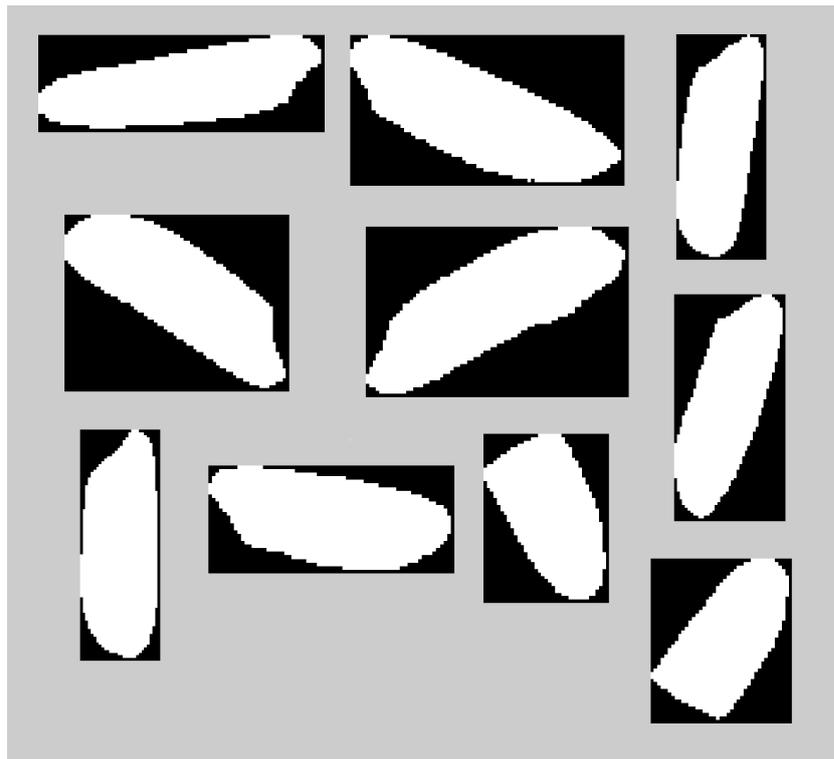


Figura 22: Imagen de los granos segmentados y etiquetado.
Elaboración: Propia.

4.4.4. Extracción de características

Los descriptores geométricos son una herramienta ampliamente utilizada en la etapa de extracción de características. Estos permiten obtener una representación numérica o matemática de las características de los objetos presentes en una imagen digital.

Los atributos geométricos son descriptores de forma comúnmente usados, a pesar de que son más simples y de que algunos de estos no son invariantes a escala, traslación y orientación. En ciertas aplicaciones han demostrado brindar

características suficientes para describir objetos y diferenciarlos correctamente a un costo computacional menor que utilizando los descriptores mencionados anteriormente. Algunos de los atributos geométricos más utilizados en el procesamiento de imágenes digitales son:

- Perímetro
- Longitud
- Área
- Centroide
- Compacidad
- Menor rectángulo que contiene al objeto (Minimum bounding box)
- Excentricidad

De entre estos atributos los que usamos son: longitud máxima y el área.

En la siguiente imagen se muestra la longitud máxima de los granos representados por una línea azul (ver figura 23):

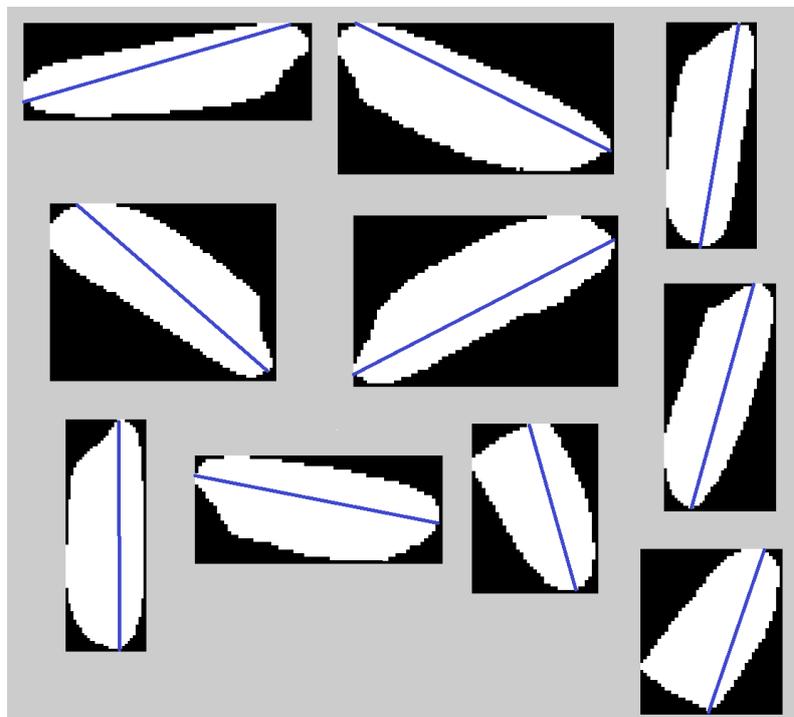


Figura 23: Imagen de los granos con la distancia máxima.
Elaboración: Propia.

En la siguiente imagen se muestra el área defectuosa del grano ubicado en la zona ventral representado por un área blanca (ver figura 24):

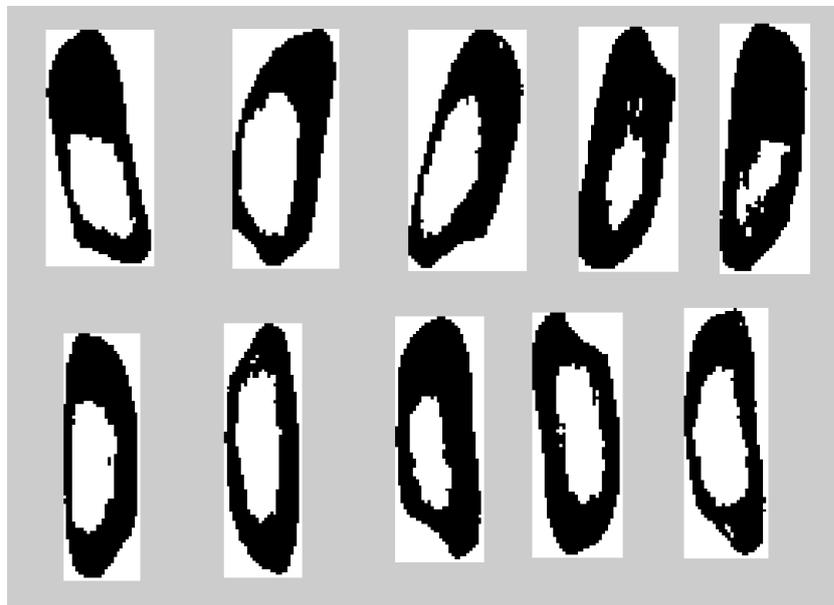


Figura 24: Imagen de los granos con el área defectuosa en el centro.
Elaboración: Propia

4.4.5. Discriminación de objetos

La discriminación de los granos se hizo con indicadores de la **Norma Técnica Peruana NTP 205.011 CEREALES Y MENESTRAS. Arroz pilado**, en donde indica claramente la clase del grano está determinada por el tamaño de los mismos como muestra la tabla siguiente (INDECOPI, 2012):

Clase	Longitud del grano de arroz pilado entero
Largo	Más de 7 mm
Mediano	De 6 mm a 7 mm
Corto	Menos de 6 mm

Tabla 10: Clase de granos de arroz pilado.
Fuente: (INDECOPI, 2012).

La **NTP 205.011** también indica para los granos tizosos:

Grano Tizosos	Área de grano en proceso de entizamiento
Tizoso Total	Más de 50 %
Tizoso Parcial	Menor al 50 %
Grano Normal	0 %

Tabla 11: Tipos de granos tizosos de arroz pilado.

Elaboración: Propia.

Fuente: (INDECOPI, 2012).

Las imágenes a continuación muestran los granos discriminados en función a su clase (ver figura 25): largo, mediano o corto con el perímetro marcado de color verde, amarillo o rojo, respectivamente.

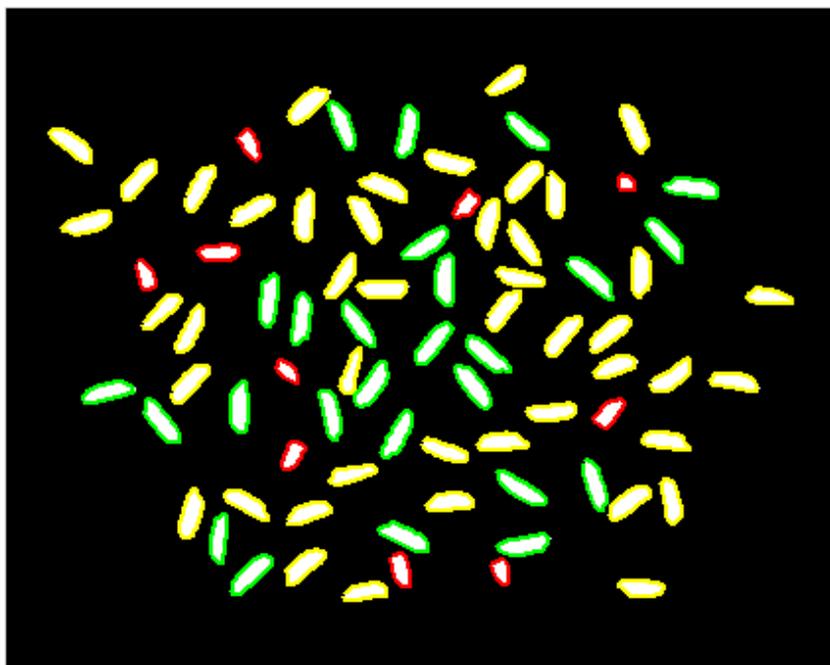


Figura 25: Imagen de los granos discriminados en función a su clase.

Elaboración: Propia.

Las imágenes a continuación muestran los granos discriminados en función a su tipo de defecto (ver figura 26): tizoso total, tizoso parcial o normal con el perímetro marcado de color rojo, amarillo o verde, respectivamente.

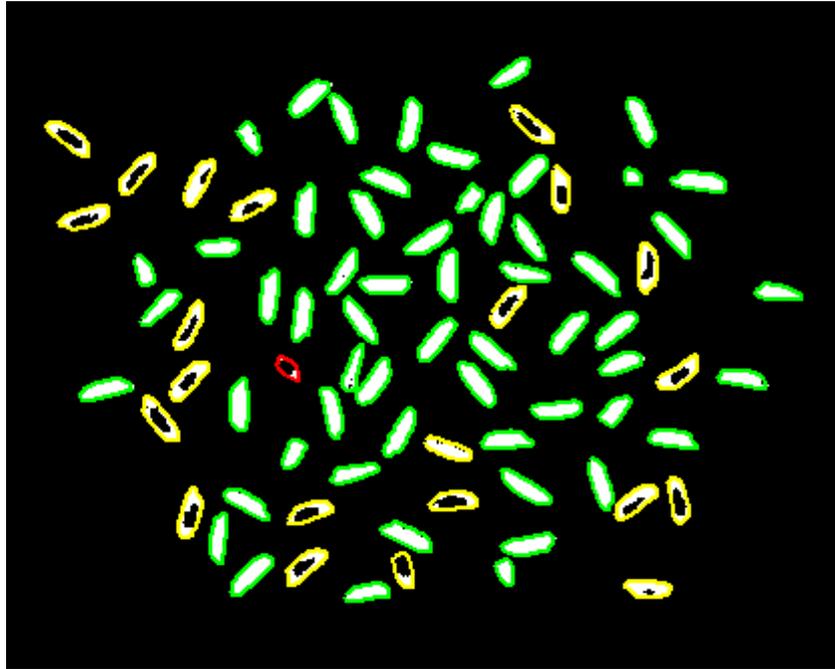


Figura 26: Imagen de los granos discriminados en función a su tipo de defecto.

Elaboración: Propia.

4.4.6. Implementación del software

La implementación del software, que comprende desde el procesamiento digital de imagen hasta la discriminación de los granos de arroz pulido, fue desarrollado en el software MatLab R2010a con el Toolbox Image Processing, bajo la metodología de desarrollo de software OpenUp.

A continuación se muestra la interfaz del software:

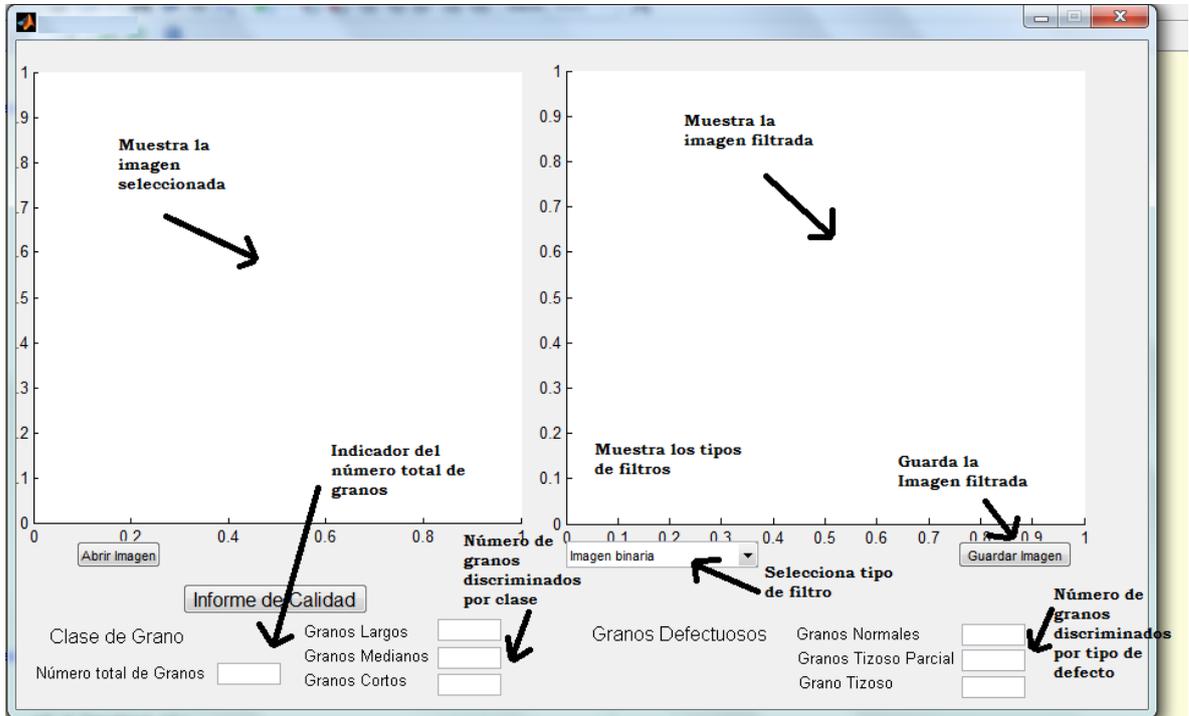


Figura 27: Interfaz gráfica del Software automático.
Elaboración: Propia.

CAPÍTULO V: RESULTADOS

5.1. PRESENTACIÓN DE RESULTADOS

5.1.1. Resultados del experimento Nro. 01

A continuación se muestran los resultados obtenidos en el Experimento Nro. 01:

GRANOS DE ARROZ PULIDO									
Método	Clase			Tipo de defecto tizoso			Total	Tiempo Empleado (Aprox.)	
	Largo	Mediano	Corto	Normal	Total	Parcial	Contabilizados	Conteo	Discriminación
Software Automático	207	156	101	448	0	16	464	20 Seg	40 Seg.
Convencional	210	160	94	445	1	18	464	12 min.	32 min.
Diferencia	3	4	7	3	1	2	0		

Tabla 12: Resultados del Experimento Nro. 01.

5.1.2. Resultados del experimento Nro. 02

A continuación se muestran los resultados obtenidos en el Experimento Nro. 02:

Clase	Número de Granos	N° Granos Errados	% Granos Errados
Largo	95	5	5.26
Mediano	87	2	2.30
Corto	33	0	0.00
Total	215	7	3.26

Tabla 13: Número de granos discriminados en función a su clase.

Tipo	Número de Granos	N° Granos Errados	% Granos Errados
Normal	188	5	2.66
Tizoso Total	5	1	20.00
Tizoso Parcial	22	2	9.09
Total	215	7	3.26

Tabla 14: Número de granos en función a su tipo de defecto.

5.1.3. Resultados del experimento Nro. 03

A continuación se muestran los resultados obtenidos en el Experimento Nro. 03:

El error absoluto está dado por:

$$\Delta x_i = |\bar{x} - x_i| \quad (5.1)$$

El error relativo está dado por:

$$Er = \frac{\Delta x_i}{\bar{x}} \quad (5.2)$$

GRANOS DE ARROZ PULIDO											
Observaciones (distancia en mm)											
Método	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	Promedio
Software Automático	7.52	7.53	7.48	7.50	7.57	7.54	7.50	7.56	7.46	7.49	7.52
Error absoluto	0.005	0.015	0.035	0.015	0.055	0.025	0.015	0.045	0.055	0.025	0.029
Error relativo	0.000665	0.001996	0.004657	0.001996	0.007319	0.003327	0.001996	0.005988	0.007319	0.003327	0.003859
$(x_i - \bar{x})^2$	0.000025	0.000225	0.001225	0.000225	0.003025	0.000625	0.000225	0.002025	0.003025	0.000625	

Tabla 15: Observaciones tomadas con el software.

$$\sum_{i=1}^N (x_i - \bar{x})^2 = 0.01125$$

La desviación estándar es la siguiente:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \tag{5.3}$$

$$\sigma = 0.03536$$

GRANOS DE ARROZ PULIDO											
Observaciones (distancia en mm)											
Método	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	Promedio
Convencional	7.45	7.40	7.35	7.45	7.40	7.40	7.45	7.30	7.35	7.45	7.40
Error absoluto	0.05	0.00	0.05	0.05	0.00	0.00	0.05	0.10	0.05	0.05	0.04
Error relativo	0.006757	0.000000	0.006757	0.006757	0.000000	0.000000	0.006757	0.013514	0.006757	0.006757	0.005405
$(x_i - \bar{x})^2$	0.002500	0.000000	0.002500	0.002500	0.000000	0.000000	0.002500	0.010000	0.002500	0.002500	

Tabla 16: Observaciones tomadas con el Vernier.

$$\sum_{i=1}^N (x_i - \bar{x})^2 = 0.025$$

La desviación estándar es la siguiente:

$$\sigma = 0.052705$$

5.2. COMPROBACIÓN DE HIPÓTESIS

5.2.1. Comprobación de las hipótesis secundarias

Hipótesis secundaria 1

H₁: El Software automático basado en procesamiento digital de imágenes reduce el tiempo en el conteo de los granos de arroz pulido.

De acuerdo a los resultados obtenidos en el experimento Nro. 01 (Ver tabla Nro. 12) el tiempo en el conteo de los granos de arroz pulido en forma convencional fue de 12 minutos, mientras que con el software automático fue de 20 segundos, lo que evidencia una reducción significativa. Por tanto la H_1 es verdadera.

Hipótesis secundaria 2

H₂: El Software automático basado en procesamiento digital de imágenes reduce el tiempo en la discriminación de los granos de arroz pulido según su clase.

De acuerdo a los resultados obtenidos en el experimento Nro. 01 (Ver tabla Nro. 12) el tiempo en la discriminación de los granos de arroz pulido según su clase en forma convencional fue de 32 minutos, mientras que con el software automático fue de 40 segundos, lo que evidencia una reducción relevante. Por tanto la H_2 es verdadera.

Hipótesis secundaria 3

H₃: El Software automático basado en procesamiento digital de imágenes reduce el tiempo en la discriminación de los granos de arroz pulido según su tipo de defecto.

De acuerdo a los resultados obtenidos en el experimento Nro. 01 (Ver tabla Nro. 12) el tiempo en la discriminación de los granos de arroz pulido según su tipo de defecto en forma convencional fue de 32, minutos mientras que con el software

automático fue de 40 segundos, lo que evidencia una reducción importante. Por tanto la H_3 es verdadera.

Hipótesis secundaria 4

H₄: El Software automático basado en procesamiento digital de imágenes reduce el margen de error en la medición de longitud de los granos de arroz pulido.

De acuerdo a los resultados obtenidos en el experimento Nro. 03 (Ver tabla Nro. 13, 14, 15 y 16) el margen de error en la medición de longitud de los granos de arroz pulido en forma convencional fue $\sigma = 0.052705$, mientras que con el software automático fue de $\sigma = 0.03536$, lo que evidencia una variación (reducción) entre ambos métodos. Por tanto la H_2 es verdadera.

5.2.2. Comprobación de la hipótesis principal

H₀: La aplicación del software automático basado en procesamiento digital de imágenes reduce el tiempo en el conteo total de granos de arroz pulido y en el proceso de discriminación por su clase y tipo de defecto, así como disminuye el margen de error en la medición de longitud.

De acuerdo a los resultados obtenidos en el experimento Nro. 01, 02 y 03 (Ver tablas Nro. 12, 13, 14, 15 y 16) que aportan evidencias significativas para concluir que la aplicación del software automático basado en procesamiento digital de imágenes reduce el tiempo en el conteo total de granos de arroz pulido y en el proceso de discriminación por su clase y tipo de defecto, así como disminuye el margen de error en la medición de longitud. Por tanto la H_0 es verdadera.

5.3. ANÁLISIS Y DISCUSIÓN DE RESULTADOS

5.3.1. Experimento Nro. 01

El experimento se trabajó con 01 (una) muestra de 10g de arroz pulido elegidos de manera aleatoria. Se realizó una observación con el método manual, y luego otra observación con el estímulo. Los resultados obtenidos muestran una gran ventaja del software en el tiempo de conteo y discriminación frente al método convencional, con una efectividad en el conteo del 100 % como nos muestra la tabla Nro. 12. Por otro lado, muestran una diferencia de conteo en el proceso de discriminación siendo 14 granos de arroz por clase y 6 granos por tipo de defecto.

5.3.2. Experimento Nro. 02

El experimento se realizó con 01 (una) muestra de 215 granos de arroz pulido, en la que la información de las características de clase y tipo de defecto se conoce de antemano. Se realizaron 03 observaciones por cada clase para verificar cuantos son los granos que por lo menos aparecen en dos clases o tipo de defecto diferentes. Los resultados (ver tabla 13) muestran que de un total de 95 granos largos 5 fueron los que aparecieron al menos una vez en la clase Mediano y Largo que vendría a ser el 5.26 % de 95 granos; de un total de 87 granos medianos fueron 2 los que aparecieron en ambas clases en distintas observaciones (Largo y Mediano) haciendo un 2.3 %. Por otro lado, los granos pequeños no manifestaron ningún error.

En la tabla Nro. 14 muestra los granos discriminados de manera errónea por el software, podemos observar que fueron 5 los granos errados en el de tipo normal, 01 grano en el de tipo tizoso total y 2 en la clase tizoso parcial. Haciendo un total de 8 granos.

5.3.1. Experimento Nro. 03

El experimento se realizó con 01 (un) grano elegido al azar y fue medido 10 (diez) veces, para determinar su margen de error en la medición. Llegando al

resultado de tener una desviación estándar de $\sigma = 0.03536$ en el método con el software automático y $\sigma = 0.052705$ con el método convencional, la desviación estándar es una medida que representa el nivel de dispersión entre los datos.

CAPÍTULO VI: CONCLUSIONES

- ❖ Se ha desarrollado e implementado en MatLab R2010a un software automático basado en procesamiento digital de imágenes para el proceso de discriminación y conteo de granos de arroz pulido por clase y tipo de defecto presentes en una imagen digital, obteniendo una confiabilidad en los resultados de 95 % para la discriminación en función a su clase, 96 % para la discriminación en función a su tipo de defecto y un 100 % para el conteo de granos.
- ❖ El algoritmo desarrollado ha permitido reducir el ruido en las siguientes imágenes digitales, esto se refleja en el alto nivel de confiabilidad obtenido en los resultados.
- ❖ Se logró reducir significativamente el tiempo que demanda el conteo de granos de arroz pulido por clase y tipo de defecto presentes en una imagen digital.
- ❖ El método empleado para el conteo y discriminación de granos de arroz pulido resultó ser simple y efectivo lo que permite al operario un entorno amigable.
- ❖ El valor umbral utilizado para la detección de áreas tizosas fue de 170 en La escala de [0 - 255].
- ❖ El error en la discriminación en función a su clase se debe a que el grano tiene un valor muy cercano a los 7mm aproximadamente 0.1 mm de margen de error y por la sensibilidad del escáner a veces lo trataba como grano largo o grano mediano. Este hecho no ocurrió en el grupo de clase corto.
- ❖ El error en la discriminación en función a su tipo de defecto se debe a que el defecto tizoso era más notorio de una parte del grano que de la otra, por eso algunos granos se contabilizaban o no como defectuoso.
- ❖ El algoritmo no se limita a alguna variedad del arroz, por el contrario puede describir muchas variedades.
- ❖ La utilización de un escáner nos permitió mantener las condiciones de luminosidad constantes.

CAPÍTULO VII: RECOMENDACIONES

Se plantean algunas recomendaciones con el fin de extender la investigación en el campo de procesamiento digital de imágenes:

- ❖ Se recomienda desarrollar un algoritmo que permita separar de manera automática los granos que se encuentran en contacto.
- ❖ Se recomienda trabajar con una misma resolución (ppp) para evitar alguna variación en los resultados.
- ❖ Para crear los ejecutables de MatLab R2010a se recomienda hacerlo en Windows x32 bits, ya que el Windows x64 carece de un compilador.

BIBLIOGRAFÍA

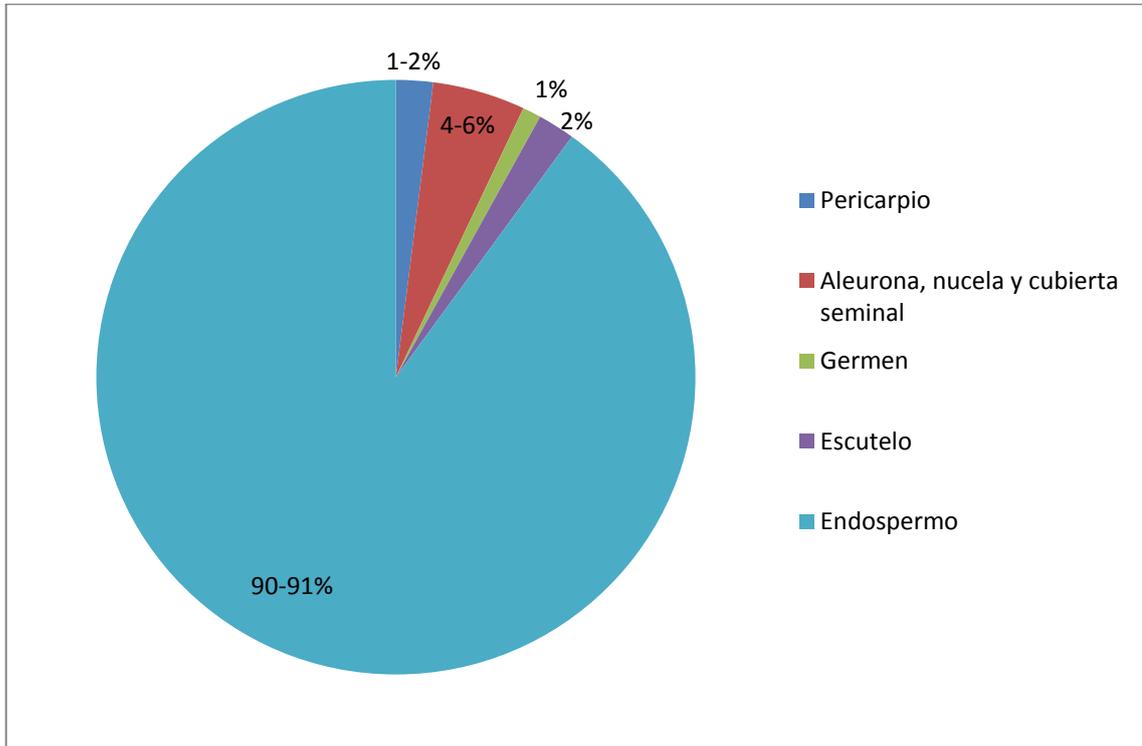
- Adobe. (s.f.). *Adobe Illustrator*. Recuperado el 18 de Noviembre de 2014, de <http://helpx.adobe.com/es/illustrator/using/importing-bitmap-images.html>
- Alexander, J., Agustín, M., & Mendoza, J. (2011). Comparación cualitativa y cuantitativa de las técnicas básicas de umbralización global basadas en histogramas para el procesamiento digital de imágenes. *Dialnet*, 266-272.
- Beltran, C. (2011). *Sistema automático de diagnóstico de parásitos intestinales a través de imágenes digitales*(Memoria tesis doctoral). Arequipa.
- Bow, S. (2002). *Pattern Recognition and Image Processing, 2nd Edition*. New York: Marcel Dekker Inc.
- Calvo, I. (s.f.). Recuperado el 12 de Noviembre de 2014, de Sitio web de. Proyectacolor: <http://www.proyectacolor.cl/aplicacion-del-color/el-color-en-el-diseno/el-color-en-pantalla/profundidad-de-color/>
- Cortés, J., Chaves, J., & Mendoza, J. (2012). Comparación cualitativa y cuantitativa de las técnicas básicas de umbralización local para el procesamiento digital de imágenes. Colombia: Redalyc Revista de investigación.
- Digitalfotored. (s.f.). <http://www.digitalfotored.com/>. Recuperado el 15 de Noviembre de 2014, de <http://www.digitalfotored.com/imagendigital/tiposimagenes.htm>
- Dirección General de Competitividad Agraria. (2013). *Principales Aspectos de la cadena Productiva del Arroz*. Lima: Publicación Ministerio de Agricultura.
- EcuRed. (s.f.). *EcuRed Enciclopedia Cubana en la Red*. Recuperado el 15 de Noviembre de 2014, de http://www.ecured.cu/index.php/EcuRed:Enciclopedia_cubana
- Esmeralda, A. L. (s.f.). Recuperado el 05 de 12 de 2012, de Sitio web de Arrocería la Esmeralda S.A.: <http://www.blanquita.com.co/>
- Gonzalez, R., & Woods, R. (2008). *Digital Image Processing, 3th edition* (Tercera ed.). States of America, United States of America: Pearson Education, Inc.
- IEEE. (1979). Nobuyuki Otsu: A threshold selection method from grey level histograms. *Institute of Electrical and Electronics Engineers.*, 62-66.
- INDECOPI. (2012). *Norma Técnica Peruana NTP 205.011 CEREALES Y MENESTRAS. Arroz Pilado*. Lima.
- INDECOPI, C. d. (31 de Octubre de 2012). NTP 205.011 Cereales y Menestras. Arroz pilado. *Norma Técnica Peruana*, 1-7.

- INEI. (Setiembre de 2014). Recuperado el 23 de Noviembre de 2014, de sitio web del Instituto Nacional de Estadística e Informática:
<http://www.inei.gob.pe/media/MenuRecursivo/boletines/00-boletin-produccion-final.pdf>
- JULIANO, O. (1994). El arroz en la nutrición humana. En *Instituto internacional de investigación sobre el arroz (FAO)* (pág. 178). Roma.
- Kiple, K. F., & Conee, K. (2000). *The Cambridge World Encyclopaedia of Food*. Cambridge, England: Cambridge University Press.
- La Serna, N., & Román, U. (2009). Técnicas de Segmentación en Procesamiento Digital de Imágenes. *Revista de Ingeniería de Sistemas e Informática vol. 6*, 9-16.
- López, J. (2009). *Análisis de imágenes microscópicas para la determinación de la cantidad y el tamaño de larvas de concha de abanico.*(Tesis de Licenciatura). Pontificia Universidad Católica del Perú. Lima.
- MERCOSUR. (s.f.). *Reglamento técnico del MERCOSUR para establecimiento de indentidad y de calidad del arroz elaborado*. Recuperado el 21 de Mayo de 2014, de Ministerio de Agricultura, Ganadería y pesca de la República Argentina:
http://64.76.123.202/SAGPYA/economias_regionales/_arroz/_publicaciones/Resoluci%F3n_SAGPYA_189-99.pdf
- Meyer, E. (s.f.). Recuperado el 15 de Octubre de 2014, de Sitio web del Grupo de Informática Aplicada al Inglés Técnico, Argentina: [http://es.tldp.org/Otros/diccionario-us-es-0.1.6/P.html#P](http://es.tldp.org/Otros/diccionario-us-es/diccionario-us-es-0.1.6/P.html#P)
- MINAG. (2009). *Producción mundial del Arroz*. Recuperado el 20 de Mayo de 2014, de Ministerio de Agricultura y Riego de la República del Perú:
<http://www.minag.gob.pe/portal/sector-agrario/agricola/cultivos-de-importancia-nacional/arroz/producci%C3%B3n>
- MINAGRI. (s.f.). *Código Alimentario Argentino*. Recuperado el 21 de Mayo de 2014, de Ministerio de Agricultura, Ganadería y pesca de la República Argentina:
http://64.76.123.202/SAGPYA/economias_regionales/_arroz/_publicaciones/CODIGO_ALIMENTARIO_ARGENTINO_Arroz.pdf
- Mustang, F. (s.f.). Recuperado el 10 de Noviembre de 2014, de Sitio web de Ford:
<http://www.fordmustang.in/>
- Olmos, S. (14 de Junio de 2014). Recuperado el 15 de Noviembre de 2014, de Asociación Correntina de Plantadores de Arroz:
http://www.acpaarrozcorrientes.org.ar/academico/Apunte_de_Calidad.pdf

- Orellana, H. (2008). *Diseño y desarrollo de un algoritmo que permita estimar el tamaño de peces, aplicando visión por computadora, y propuesta realizar la selección adecuada de dichos peces (Tesis de titulación)*. Pontificia Universidad Católica del Perú. Lima.
- Pérez, I., & Maria, M. (2009). Calidad del grano y variabilidad genética de variedades y líneas de arroz del Instituto nacional de investigaciones agrícolas (INIA). *Agronomía Tropical*, 445-456.
- ProdSTAT. (2013). *Organización de las Naciones Unidas para la Alimentos y la Agricultura*. Recuperado el 04 de Noviembre de 2014, de <http://faostat3.fao.org/>
- RAE. (s.f.). Recuperado el 14 de Noviembre de 2014, de Real Academia Española: <http://lema.rae.es/dpd/?key=p%C3%ADxel>
- Romo, H., Ramírez, F., & Valdivieso, R. (2007). Detección del Bacilo Mycobacterium Tuberculosis mediante Reconocimiento de Patrones. *Redalyc*, 111-114.
- Sampallo, G. M., Acosta, C., Cleva, M., & González, A. (06 de 29 de 2013). *Determinación de las características de aspecto de granos de arroz pulidos empleando procesamiento de imágenes*. Recuperado el 03 de 05 de 2014, de Universidad Tecnológica Nacional: <http://www.frre.utn.edu.ar/IJCYT/clean/files/get/item/2201>
- Sandoval, A., Rodriguez, E., & Fernandez, A. (2005). Aplicación del análisis por calorimetría diferencial de barrido (DSC) para la caracterización de las modificaciones del almidón. *SciELO*, 45-53.
- Sato, J. (2013). *Microscopio Automatizado: Conteo de bacilos de tuberculosis (Tesis para optar título)*. Pontificia Universidad Católica del Perú. Lima.
- SEDESOL, S. d. (2014). *Instructivo para el muestreo y análisis de maíz, frijol y arroz*. Mexico D.F.
- Shutler, J. (s.f.). Recuperado el 10 de Noviembre de 2014, de Sitio web de The University of Edinburgh: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/SHUTLER3/CVonline_moments.html
- Theodoridis, S., & Koutroumbas, K. (2009). *Pattern Recognition, 4th Edition*. New York: Academic Press.
- Tkalcic, M. T. (2003). En *Colour spaces - perceptual, historical and applicational background* (págs. 304-308). Liubliana, Eslovenia.
- Universidad Nacional de Quilmes. (Octubre de 2005). *Ingeniería en Automatización y Control Industrial*. Recuperado el 14 de Octubre de 2014, de Página web de la Universidad Nacional de Quilmes: <http://iaci.unq.edu.ar/>

ANEXOS

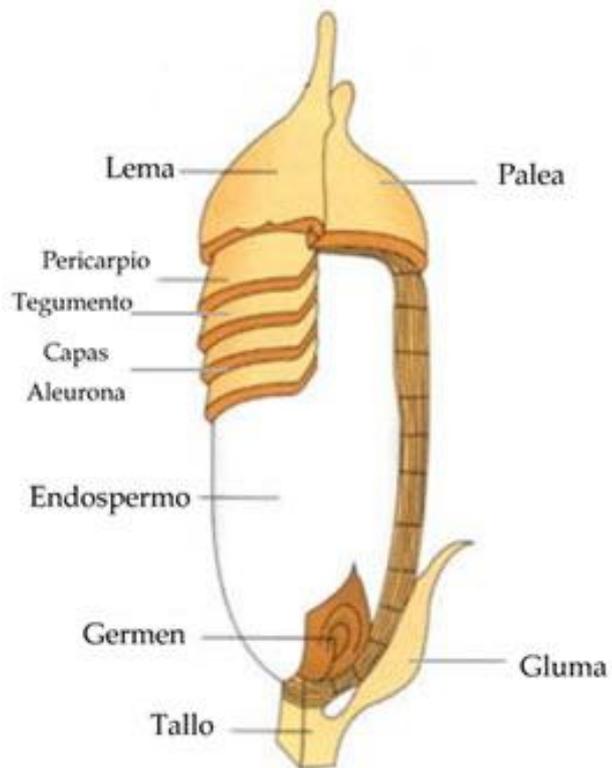
ANEXO 1: Distribución de peso del arroz



Fuente: (Esmeralda, 2012).

Elaboración: Propia.

ANEXO 2: Partes de un grano de arroz



Fuente: (Esmeralda, 2012).

Elaboración: Propia.

ANEXO 3: Operacionalización de variables.

Variables	Definición conceptual	Definición operacional	Indicadores	Unidad
X: Software automático basado en procesamiento digital de imágenes	Es un algoritmo desarrollado para resaltar las características relevantes en las imágenes digitales.	El algoritmo de procesamiento digital de imágenes será desarrollado bajo la metodología de desarrollo de software OpenUp.	Ausencia / Presencia	
Y1: Tiempo en el conteo de los granos de arroz pulido.	Es el tiempo de demora en contar el número total de granos presentes en una imagen.	El tiempo se medirá empleando un cronómetro	Tiempo	Minutos / Segundos
Y2: Tiempo en la discriminación de los granos de arroz pulido según su clase.	Es el tiempo que demanda discriminar los granos de arroz pulido según su clase.	El tiempo se medirá empleando un cronómetro	Tiempo	Minutos / Segundos
Y3: Tiempo en la discriminación de los granos de arroz pulido según su tipo de defecto.	Es el tiempo que demanda discriminar los granos de arroz pulido según su tipo de defecto.	El tiempo se medirá empleando un cronómetro	Tiempo	Minutos / Segundos
Y4: Margen de error en la medición de los granos de arroz pulido.	Es el grado de dispersión que tiene una cierta cantidad de medidas.	Formula de desviación estándar.	Números reales positivos	

ANEXO 4: Matriz de consistencia.

TÍTULO	PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLES
Software automático de reconocimiento de patrones para la discriminación de granos de arroz pulido basado en procesamiento digital de imágenes	PROBLEMA GENERAL ¿Cuál es el efecto al aplicar un sistema automático basado en procesamiento digital de imágenes en la determinación del número total de granos de arroz pulido en una imagen digital y su discriminación en función a su clase y tipo de defecto?	OBJETIVO GENERAL Implementar un sistema capaz de determinar el número total de granos de arroz pulido en una imagen digital y discriminarlos en función a su clase y tipo de defecto, de manera automática utilizando procesamiento digital de imágenes.	HIPÓTESIS GENERAL La aplicación del software automático basado en procesamiento digital de imágenes reduce el tiempo en el conteo total de granos de arroz pulido y en el proceso de discriminación por su clase y tipo de defecto, así como disminuye el margen de error en la medición de longitud.	VARIABLE INDEPENDIENTE X: Software automático basado en procesamiento digital de imágenes VARIABLE DEPENDIENTE Y1: Tiempo en el conteo de los granos de arroz pulido. Y2: Tiempo en la discriminación de los granos de arroz pulido según su clase. Y3: Tiempo en la discriminación de los granos de arroz pulido según su tipo de defecto. Y4: Margen de error en la medición de los granos de arroz pulido.
	Problema específico 1 ¿Cómo el Software automático basado en procesamiento digital de imágenes afecta al tiempo en el conteo de los granos de arroz pulido?	Objetivo específico 1 Implementar un algoritmo que nos permita determinar el número total de granos en una imagen digital basándose en el procesamiento digital de imágenes.	Hipótesis secundaria 1 El Software automático basado en procesamiento digital de imágenes reduce el tiempo en el conteo de los granos de arroz pulido.	
	Problema específico 2 ¿Cómo el Software automático basado en procesamiento digital de imágenes afecta al tiempo en la discriminación de los granos de arroz pulido según su clase?	Objetivo específico 2 Diseñar un algoritmo que pueda discriminar los granos de arroz pulido en función a su clase (tamaño largo, mediano o corto) basándose en el procesamiento digital de imágenes.	Hipótesis secundaria 2 El Software automático basado en procesamiento digital de imágenes reduce el tiempo en la discriminación de los granos de arroz pulido según su clase.	

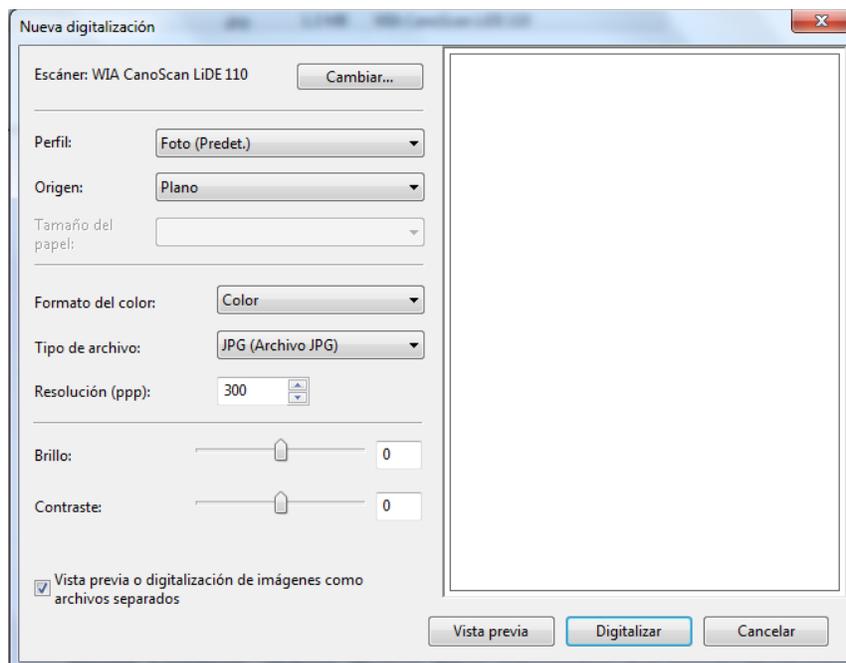
	<p>Problema específico 3</p> <p>¿Cómo el Software automático basado en procesamiento digital de imágenes afecta al tiempo en la discriminación de los granos de arroz pulido según su tipo de defecto?</p>	<p>Objetivo específico 3</p> <p>Diseñar un algoritmo que pueda discriminar los granos de arroz pulido en función a su tipo de defecto (grano normal, tizoso total o tizoso parcial) basándose en el procesamiento digital de imágenes.</p>	<p>Hipótesis secundaria 4</p> <p>El Software automático basado en procesamiento digital de imágenes reduce el margen de error en la medición de longitud de los granos de arroz pulido.</p>	
	<p>Problema específico 4</p> <p>¿De qué manera el Software automático basado en procesamiento digital de imágenes se relaciona con margen de error en la medición de los granos de arroz pulido?</p>	<p>Objetivo específico 4</p> <p>Diseñar un algoritmo que nos permita reducir el ruido en las imágenes digitales y resaltar las características de los granos para una mejor interpretación de datos.</p>		
		<p>Objetivo específico 5</p> <p>Implementar en el software en MatLab R2010a los algoritmos para determinar el número total de granos, realizar la discriminación de los granos de arroz pulido en función a su clase y tipo de defecto, que permita reducir el ruido en las imágenes digitales, resaltar las características de los granos y reducir el margen de error en la medición de longitud.</p>		

ANEXO 5: Manual de Usuario

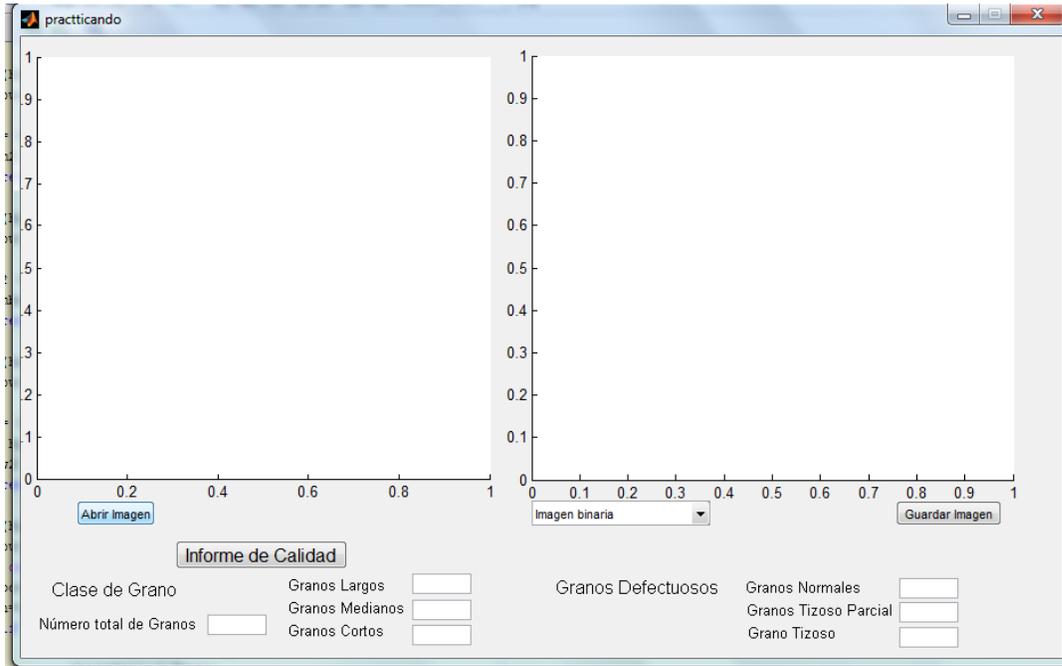
Paso 1: Colocar los granos de arroz pulido sobre el escáner



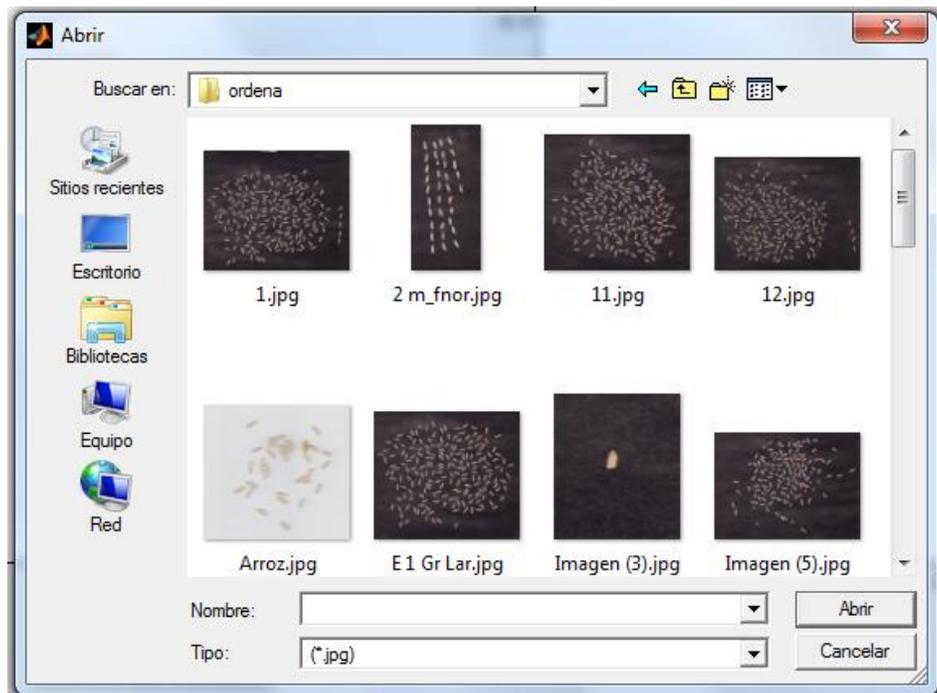
Paso 2: Haciendo uso del software del escáner, iniciar la digitalización de la imagen (300ppp).



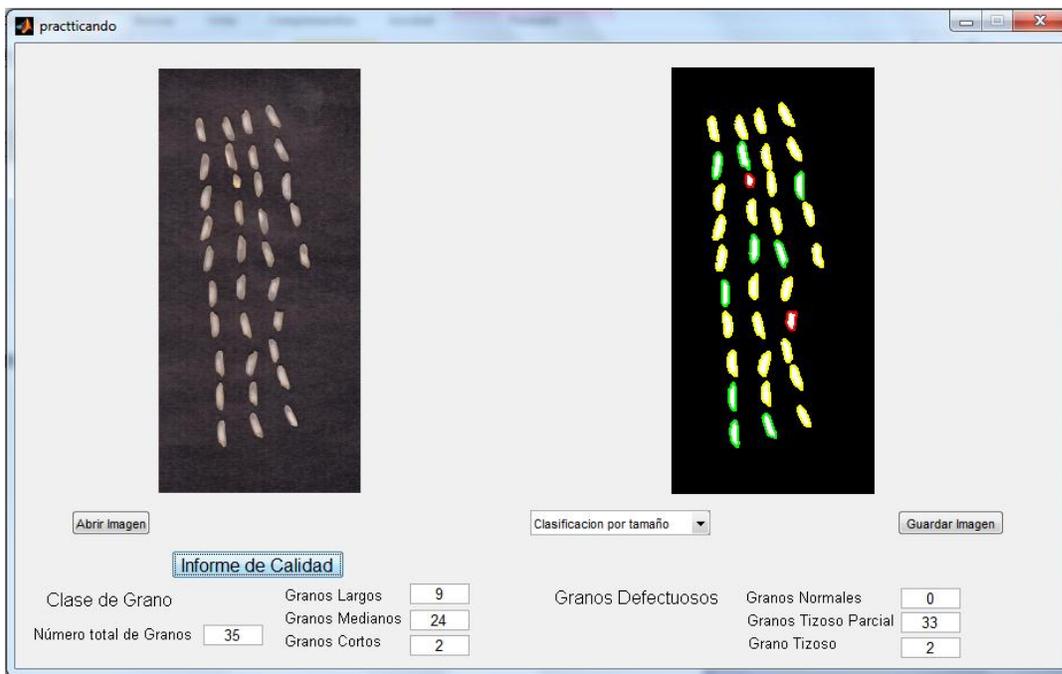
Paso 3: Iniciar el Software automático.



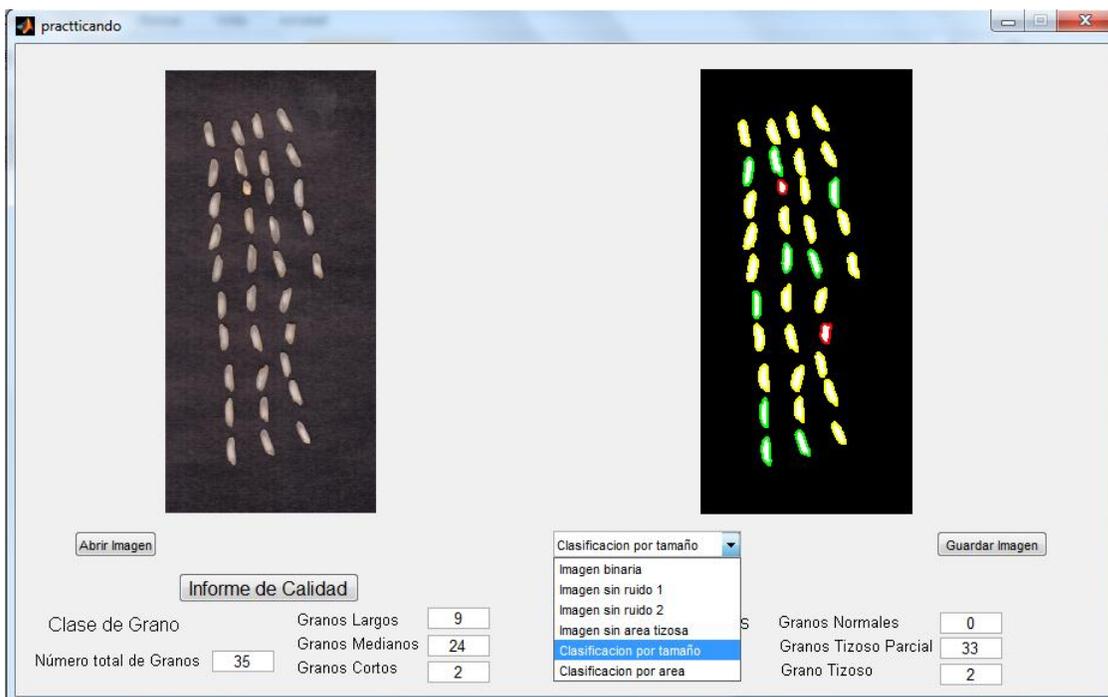
Paso 4: Clic en “Abrir imagen” y seleccionar la imagen previamente escaneada.



Paso 5: Clic en “Informe de calidad” para obtener el número de granos discriminados por clase y tipo de defecto.



Paso 6: Existen distintos tipos de filtros desarrollados.



ANEXO 6: Imágenes RGB de un grano de arroz con el área tizosa

Vista de Grano Normal



Vista de Grano con área tizosa



Porcentaje del área tizosa de un grano calculada por el software: 0.3504.

Vista de Grano Normal



Vista de Grano con área tizosa



Porcentaje del área tizosa de un grano calculada por el software: 0.3509

Vista de Grano Normal



Vista de Grano con área tizosa



Porcentaje del área tizosa de un grano calculada por el software: 0.3808

Vista de Grano Normal



Vista de Grano con área tizosa



Porcentaje del área tizosa de un grano calculada por el software: 0.3396