

UNIVERSIDAD MAYOR DE SAN ANDRÉS

FACULTAD DE CIENCIAS PURAS Y NATURALES

CARRERA DE INFORMÁTICA



TESIS DE GRADO

**SEGMENTACIÓN DE IMÁGENES DE CROMATOGRAFÍA
DE SUELOS BASADO EN EL ALGORITMO K-MEANS**

Para obtener el Título de Licenciatura en Informática

Mención Ingeniería de Sistemas Informáticos

POSTULANTE: KATERINE ANDREA RAMALLO UNZUETA

TUTOR METODOLÓGICO: M.SC. ALDO RAMIRO VALDEZ ALVARADO

ASESORA: M.SC. LUISA VELÁSQUEZ LÓPEZ

LA PAZ – BOLIVIA

2016



**UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE INFORMÁTICA**



LA CARRERA DE INFORMÁTICA DE LA FACULTAD DE CIENCIAS PURAS Y NATURALES PERTENECIENTE A LA UNIVERSIDAD MAYOR DE SAN ANDRÉS AUTORIZA EL USO DE LA INFORMACIÓN CONTENIDA EN ESTE DOCUMENTO SI LOS PROPÓSITOS SON ESTRICTAMENTE ACADÉMICOS.

LICENCIA DE USO

El usuario está autorizado a:

- a) visualizar el documento mediante el uso de un ordenador o dispositivo móvil.
- b) copiar, almacenar o imprimir si ha de ser de uso exclusivamente personal y privado.
- c) copiar textualmente parte(s) de su contenido mencionando la fuente y/o haciendo la referencia correspondiente respetando normas de redacción e investigación.

El usuario no puede publicar, distribuir o realizar emisión o exhibición alguna de este material, sin la autorización correspondiente.

TODOS LOS DERECHOS RESERVADOS. EL USO NO AUTORIZADO DE LOS CONTENIDOS PUBLICADOS EN ESTE SITIO DERIVARA EN EL INICIO DE ACCIONES LEGALES CONTEMPLADOS EN LA LEY DE DERECHOS DE AUTOR.

DEDICATORIA

A mis padres, por todo el esfuerzo que hicieron para darme la oportunidad y haberme brindado todos los medios para realizarme profesionalmente.

A mis hermanas, porque fueron mi motivación para seguir adelante y de igual manera quiero ser una motivación para ellas.

AGRADECIMIENTOS

Principalmente a mi madre Vivian Unzueta, mi agradecimiento hacia ti siempre será infinito, has sido un ejemplo de perseverancia para lograr mis objetivos, además de haberme dado todo lo necesario para formarme profesionalmente y como persona.

A mis hermanas Camila, Joseline y Nathalie, por sus palabras de aliento en momentos difíciles, por su apoyo incondicional y por su cariño, por nunca haber perdido la confianza en mí.

A mi padre Fernando Ramallo, por haberme formado, por todas sus enseñanzas a través de toda mi vida y por sus palabras que han forjado mi carácter.

A mi tutor M. Sc. Aldo Ramiro Valdez Alvarado, por su tiempo, paciencia, dedicación y desprendimiento al momento de transmitir sus conocimientos, que han sido una base fundamental para la conclusión de este trabajo.

A mi asesora M Sc. Luisa Velásquez López, por su tiempo y dedicación para guiarme, en la elaboración de este trabajo de investigación, por sus consejos y su apoyo tanto en el ámbito profesional, como personal.

A María Condori por apoyarme y aconsejarme en esta etapa, por su ayuda y motivación para concluir este trabajo de investigación.

A mis amigas Lisette Nina y Camila Delgado por su amistad, cariño, consejos y apoyo durante el proceso de elaboración de este trabajo de investigación.

RESUMEN

El procesamiento de imágenes digitales (PDI) nos permite la extracción de información de éstas, y dentro de esta disciplina la segmentación permite la identificación de las partes que lo constituyen. La segmentación de imágenes tiene aplicaciones en reconocimiento de patrones y sistemas de control de tráfico entre otros.

La aplicación de la segmentación de imágenes digitales mediante técnicas de agrupamiento o no supervisadas, ha dado buenos resultados en la colaboración para el diagnóstico de ciertas enfermedades a través de imágenes médicas digitales. De manera análoga en el presente trabajo se plantea segmentación de imágenes de cromatografía de suelo, con el fin de colaborar en el diagnóstico del estado de los mismos.

Existen distintas técnicas de segmentación de imágenes, en el presente trabajo se plantea un procedimiento de segmentación de imágenes de cromatografía de suelos basado en el algoritmo k-means. Este algoritmo permite dividir un conjunto de datos de entrada en subconjuntos (clusters), de tal manera que los elementos de cada subconjunto compartan cierto patrón, una o varias características a priori desconocidas.

Este procedimiento, fue aplicado buscando la extracción de los segmentos correspondientes a las zonas central, interna, intermedia, externa y periférica que constituyen la estructura de un cromatograma, obteniendo los resultados esperados.

En el presente trabajo también se plantea el prototipo de una aplicación móvil con el cual se pretende colaborar en el proceso de diagnóstico del estado de los suelos implementado el algoritmo planteado.

ABSTRACT

The digital image processing (PDI) allows us to extract information from them, and within this discipline the segmentation allows the identification of the parts that constitute it. Image segmentation has applications in pattern recognition and traffic control systems, among others.

The application of segmentation of digital images using grouping techniques or unsupervised, has given good results in the collaboration for the diagnosis of certain diseases through digital medical images. Similarly, the present work proposes the segmentation of images of soil chromatography, in order to collaborate in the diagnosis of the state of the same.

There are different techniques of segmentation of images, in the present work we propose a procedure of segmentation of images of soil chromatography based on the k-means algorithm. This algorithm allows dividing a set of input data into subsets, so that the elements of each subset share a certain pattern, one or more unknown a priori characteristics.

This procedure was applied to extract the segments corresponding to the central, internal, intermediate, external and peripheral areas that constitute the structure of a chromatogram, obtaining the expected results.

In the present work, also the prototype of a mobile application is proposed with which it is tried to collaborate in the process of diagnosis of the state of the ground, implemented the proposed algorithm.

ÍNDICE

CAPÍTULO 1	1
MARCO INTRODUCTORIO.....	1
1.1 INTRODUCCIÓN.....	1
1.2 ANTECEDENTES	2
1.3 PLANTEAMIENTO DEL PROBLEMA	4
1.3.1 PROBLEMA CENTRAL	5
1.3.2 PROBLEMAS SECUNDARIOS	6
1.4 DEFINICIÓN DE OBJETIVOS.....	6
1.4.1 OBJETIVO GENERAL.....	6
1.4.2 OBJETIVOS ESPECÍFICOS	7
1.5 HIPÓTESIS	7
1.6 OPERACIONALIZACIÓN DE VARIABLES	7
1.7 JUSTIFICACIÓN.....	7
1.7.1 JUSTIFICACIÓN ECONÓMICA.....	7
1.7.2 JUSTIFICACIÓN SOCIAL.....	8
1.7.3 JUSTIFICACIÓN CIENTÍFICA.....	8
1.8 ALCANCES Y LÍMITES	8
1.8.1 ALCANCES	9
1.8.2 LÍMITES.....	9
1.9 APORTES	9
1.9.1 PRÁCTICO.....	9

1.9.2	TEÓRICO	10
1.10	METODOLOGÍA.....	10
	CAPÍTULO 2	11
	MARCO TEÓRICO	11
2.1	IMAGEN DIGITAL	11
2.1.1	CONCEPTOS BÁSICOS DE IMAGEN DIGITAL.....	11
2.1.2	REPRESENTACIÓN DE IMÁGENES	12
2.1.3	HISTOGRAMA.....	14
2.2	ÁREAS RELACIONADAS CON IMÁGENES	15
2.3	PROCESAMIENTO DIGITAL DE IMÁGENES	16
2.3.1	PROCESAMIENTO DE BAJO NIVEL	18
2.3.2	PROCESAMIENTO DE NIVEL MEDIO	18
2.3.3	PROCESAMIENTO DE ALTO NIVEL.....	19
2.4	SEGMENTACIÓN DE IMÁGENES.....	20
2.5	TÉCNICAS DE RECONOCIMIENTO DE PATRONES	21
2.6	ALGORITMO K-MEANS	22
2.7	CROMATOGRAFÍA DE SUELOS	24
2.8	INTERPRETACIÓN DE CROMATOGRAMAS.....	25
2.8.1	ZONAS DEL CROMATOGRAMA	25
2.8.2	COLORACIÓN DEL CROMATOGRAMA	27
2.8.3	INTEGRACIÓN DE ZONAS	28
2.8.4	FORMAS	28

2.8.5	PATRONES.....	30
	CAPÍTULO 3	32
	MARCO APLICATIVO	32
3.1	INTRODUCCIÓN.....	32
3.2	DESCRIPCIÓN DEL MÉTODO	33
3.2.1	ADQUISICIÓN DE IMÁGENES	34
3.2.2	PRE-PROCESAMIENTO	36
3.2.3	HISTOGRAMA Y DESCRIPTORES.....	36
3.3	SEGMENTACIÓN.....	38
3.3.1	EXTRACCIÓN DEL ÁREA DE INTERÉS	38
3.3.2	SEGMENTACIÓN DEL ÁREA DE INTERÉS	39
3.4	PLANTEAMIENTO DEL ALGORITMO.....	40
3.5	EVALUACIÓN DE LA SEGMENTACIÓN.....	46
3.6	REPRESENTACIÓN	49
3.7	RECONOCIMIENTO	50
3.8	DISEÑO DEL PROTOTIPO	51
3.8.1	DECLARACIÓN DE PROPÓSITOS	51
3.8.2	LISTA DE ACONTECIMIENTOS.....	51
3.8.3	DESCRIPCIÓN DE PROCESOS.....	51
3.8.4	DIAGRAMAS DE FLUJO DE DATOS	55
3.8.5	IMPLEMENTACIÓN DE ALGORITMOS.....	57
3.8.6	INTERFAZ DEL PROTOTIPO	60

CAPITULO 4	61
PRUEBA DE HIPÓTESIS	61
4.1 INTRODUCCIÓN	61
4.2 PRUEBA DE RACHAS DE WALD-WOLFOWITZ	61
4.3 ESTADO DE LA HIPÓTESIS	63
CAPITULO 5	67
CONCLUSIONES Y RECOMENDACIONES	67
5.1 INTRODUCCIÓN	67
5.2 CONCLUSIONES	67
5.3 RECOMENDACIONES	68
BIBLIOGRAFÍA	69

ÍNDICE DE FIGURAS

figura 2.1 Gama de áreas relacionadas con el tratamiento de imágenes	16
figura 2.2. Etapas del Procesamiento Digital de Imágenes	17
figura 3.1 Diagrama de secuencia del procesamiento digital de imágenes de Cromatografía de suelos	33
figura 3.2 Diagrama de bloques del método propuesto	34
figura 3.3 Cromatograma con su respectivo histograma	37
figura 3.4 Secuencia de procesos de la segmentación	38
figura 3.5 Imagen en escala de grises (izquierda); imagen umbralizada (derecha).....	39
figura 3.6 Entrada y salida de la representación de imagen	40
figura 3.7 Diagrama de Flujo cálculo de descriptores	41
figura 3.8 Diagrama de flujo del proceso de División por niveles	41
figura 3.9 Diagrama de flujo Ajuste de Histograma.....	43
figura 3.10 Diagrama de flujo del proceso de umbralización.....	44
figura 3.11 Diagrama de flujo de la segmentación mediante K-means.....	45
figura 3.12 Cromatogramas de dos diferentes tipos de suelo	46
figura 3.13 Componentes de la segmentación de un cromatograma	50
figura 3.14 Diagrama de Entrada-Proceso-Salida del cálculo de descriptores.....	52
Figura 3.15 Diagrama Entrada-Proceso- Salida de la división por niveles	53
figura 3.16 Diagrama Entrada-Proceso- Salida de la división por niveles.....	53
Figura 3.17 Diagrama Entrada-Proceso- Salida de la umbralización.....	54
figura 3.18 Diagrama Entrada-Proceso- Salida de la segmentación	55
figura 3.19 Diagrama en bloques del proceso de reconocimiento de patrones	55

figura 3.20 Prototipo de aplicación de procesamiento de imágenes de cromatogramas.	56
figura 3.21 Prototipo de aplicación de procesamiento de imágenes de cromatogramas. Diagrama de primer nivel.....	56

ÍNDICE DE TABLAS

Tabla 1.1 Áreas degradadas por municipio de la Zona de los Yungas - La Paz	5
Tabla 3.1 Tabla de valores para la delimitación entre valores oscuros, medios y claros	36
Tabla 3.2 Resultados de segmentación con algoritmo k-means para distintos valores de k	48
Tabla 3.3 Comparación de segmentos adquiridos según el valor de k.....	49
Tabla 4.1 Muestras tomadas para evaluación del prototipo	64
Tabla 4.2 Prueba de Rachas.....	64

MARCO INTRODUCTORIO

1.1 INTRODUCCIÓN

En la actualidad el campo de la Agricultura se encuentra en su tercera revolución, la cual está siendo liderada por la biotecnología y la informática, de tal manera que los productores agrícolas pueden tener acceso rápido y preciso a la información, haciendo que el proceso de planificación de sus actividades en cuanto a producción y el posicionamiento de sus productos en el mercado sea mucho más eficiente, eficaz y competitivo.

Una de las problemáticas más importantes de la Agricultura es el análisis de los suelos para el diagnóstico del estado de los mismos, este proceso es indispensable para determinar la usabilidad y la fertilidad de estos, sin embargo, la informática y la biotecnología aún no han abordado esta problemática.

Las características físico-químicas del suelo, deben ser conocidas por el productor agrícola, ya que el crecimiento y desarrollo de los cultivos y la cantidad y calidad de las cosechas, están en relación directa con los nutrimentos y las características de los suelos.

Una de las técnicas más utilizadas, para el análisis de los suelos es la Cromatografía por ser el método más económico y relativamente sencillo de aplicar, de esta técnica se obtiene una imagen en papel llamada cromatograma a través del cual se hace el diagnóstico del estado de los suelos identificando ciertos patrones y colores.

El uso de algoritmos de agrupamiento (en inglés, *clustering*) en distintos ámbitos como la medicina para la identificación de patrones en imágenes han tenido buenos resultados, brindando mayor precisión al momento de dar un diagnóstico en algunas enfermedades a través de imágenes digitales, del mismo modo, podrían ser de utilidad en el ámbito de la

agricultura particularmente para el diagnóstico del estado de los suelos a través de imágenes digitales de cromatogramas.

El presente trabajo se plantea el uso del algoritmo de Agrupamiento K-means para la segmentación de imágenes de cromatografía de suelos cuyo objetivo principal es el de brindar un diagnóstico preciso del estado de los mismos de esta manera se busca apoyar en la gestión sostenible de manejo de los suelos.

1.2 ANTECEDENTES

A partir de los años sesenta y con el desarrollo de las computadoras, el procesamiento digital de imágenes ha avanzado a pasos impresionantes y sus algoritmos se empezaron a utilizar en un amplio rango de aplicaciones (Geólogos que estudian la contaminación con imágenes satelitales o aéreas, arqueólogos que restauran viejas fotografías de antiguos artefactos destruidos con el tiempo aplicaciones en astronomía, biología, defensa, la industria, reconocimiento de caracteres, control de calidad, entre otros).

En los años 60's la NASA uso el procesamiento de imágenes para el reconocimiento y caracterización de la superficie lunar en su Programa "APOLO" y también para procesar imágenes digitales de planetas como Marte Venus y Mercurio en su Proyecto "Mariner". Por otra parte, en el área de la medicina, se empezó a utilizar el Procesamiento de imágenes en imágenes de Rayos X.

Ya en los años 70's se empezaron a procesar imágenes digitales de cultivos agrícolas e imágenes meteorológicas, también en el área de la biología para la detección y clasificación automática de células.

Entre los años 80 y 90 se empezó a utilizar el procesamiento de imágenes para efectos especiales en la televisión, en el área de medicina para TAC, Resonancia magnética y ultra sonidos entre otros, y en la industria para el control de calidad.

Actualmente el procesamiento de imágenes es usado para mejoras en imágenes, visión artificial, diagnóstico asistido por computador entre otros usos.

Entre los trabajos que por su contenido constituyen una base de antecedentes para la elaboración del presente trabajo, por similares en cuanto a la metodología empleada para el tratamiento de imágenes digitales, se tienen los siguientes:

- **Análisis de la aplicación de algoritmos de K-means y Continuous Max-Flow a la segmentación de imágenes en color** (Pinto L., 2015).

En el Departamento de Teoría de la Señal y Comunicaciones Escuela Técnica Superior de Ingeniería de la Universidad de Sevilla se hizo un análisis del uso de algoritmos k-means y Continuous Max-Flow para la segmentación de imágenes en color que describe la precisión que tienen estos algoritmos para identificar patrones, bordes y regiones en imágenes en color.

- **Segmentación de imágenes a color mediante técnicas de agrupamiento de datos empleando los algoritmos k-means y c-means** (Guevara G., 2015).

En la Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas-IPN. México, D. F. se publicó un artículo científico el cual describe la diferencia de precisión en la segmentación de imágenes entre el algoritmo k-means y c-means.

- **Sistema de diagnóstico de la retinopatía diabética mediante imágenes digitales** (Flores, 2014).

En la Carrera de Informática de la Universidad de San Andrés se desarrolló el Sistema de diagnóstico de retinopatía diabética mediante imágenes digitales el cual se basa en segmentación para el procesamiento de las imágenes.

- **Implementación del algoritmo de Fourier para identificación de especies vegetales adoptándolo a dispositivos móviles Android.** (Herrera, 2014)

En la Carrera de Informática de la Universidad de San Andrés se desarrolló la aplicación móvil para la identificación de especies vegetales usando técnicas de procesamiento de imágenes mediante la implementación de algoritmo de Fourier.

- **Cuantificación y detección de alteraciones morfológicas de glóbulos rojos en imágenes digitales microscópicas (Tarqui, 2014)**

En la Carrera de Informática de la Universidad de San Andrés se desarrolló el Sistema para a cuantificación y detección de alteraciones morfológicas en glóbulos rojos que usa técnicas de procesamiento de imágenes y redes neuronales.

1.3 PLANTEAMIENTO DEL PROBLEMA

En julio de 2015, Ronald Vargas, Secretario de la Alianza Mundial por el Suelo FAO¹ y expositor del Simposio “Manejo Sostenible de Suelos”, indicó que el 50% de tierras agrícolas y pecuarias de Bolivia se encuentran en estado de degradación lo que implica erosión de suelos, pérdidas de la biodiversidad y de nutrientes del suelo, entre otras.

Particularmente en la zona de Los Yungas, Norte de La Paz, los monocultivos, especialmente de coca están degradando los suelos en esta región, según El Estudio de Degradación de los Suelos de Los Yungas de La Paz, realizado por ECOTOP² con el apoyo del FONADAL³ (2007-2008), los niveles o grados de degradación de suelos obtenidos en base a análisis de los mismos y la interpretación de las imágenes de satélite por municipio, muestran los siguientes resultados:

MUNICIPIO	ÁREA DE DEGRADACIÓN (ha)			ÁREA TOTAL MUNICIPIO (ha)
	ALTO	MEDIO	BAJO	
Cajuata	11.267,35	29.367,63	29.174,79	69.809,76
Chulumani	10.586,69	5.745,92	12.143,52	28.476,13
Coripata	9.582,49	4.854,76	54.609,48	69.046,73
Coroico	7.498,22	7.695,52	90.429,84	105.623,58

¹ FAO: Food and Agriculture Organization of the United Nations (Organización de las Naciones Unidas para

² ECOTOP: Empresa Consultora en desarrollo rural y agricultura ecológica

³ FONADAL: Fondo Nacional de Desarrollo Alternativo

Irupana	24.032,14	55.973,61	55.955,29	135.961,04
Yanacachi	25.134,04	4.552,27	28.530,95	58.217,25
TOTAL	88100,94	108189,7	270843,86	467134,49

Tabla 1.1 Áreas degradadas por municipio de la Zona de los Yungas - La Paz
Fuente: Instituto Nacional de Estadísticas Bolivia

Esto por realizar prácticas insostenibles, como el uso excesivo de maquinaria agrícola en el occidente y oriente del país, el uso de fertilizantes químicos porque el suelo está perdiendo la capacidad de producir, la falta de normativas que regulen el uso y manejo del recurso suelo, la deforestación en áreas no permitidas y la quema de pastizales, cultivos y restos de deforestación. Esto amenaza la futura seguridad alimentaria y los medios de sustento de los pequeños, medianos y grandes agricultores.

Por otra parte, en la Zona de Alto Beni (La Paz), cuya producción agrícola es principalmente el Cacao, se observa que existe una diferencia significativa entre las hectáreas de producción entre las áreas en las que se divide este territorio, con lo que se infiere que la siembra de este producto, no se realiza de forma planificada ni se hace un seguimiento homogéneo del uso y administración del suelo; sino más bien se realiza con decisión particular de los agricultores y sus necesidades (Baldivieso y Zapata, 2006).

Uno de los primeros pasos para hacer un manejo sostenible de tierras es realizar un estudio de suelos y un análisis de laboratorio. Sin embargo, gran parte de las tierras han sido utilizadas sin estudios previos que muestren cual es el tipo de uso más adecuado y cuál es el efecto ambiental de los diferentes usos. (Ugarte, 2015)

1.3.1 PROBLEMA CENTRAL

A partir de la anterior problemática se plantea el siguiente problema:

- ¿Cómo apoyar en el diagnóstico del estado de los suelos?

1.3.2 PROBLEMAS SECUNDARIOS

A partir de la problemática anteriormente planteada, se tienen los siguientes problemas secundarios:

- Los agricultores no realizan un análisis previo del estado de los suelos para conocer el uso más adecuado de estos, lo cual ocasiona el rápido deterioro de los mismos.
- Muchos tipos de uso de la tierra son hechos de forma y en lugares inadecuados, lo que ha resultado en pobreza, degradación ambiental, explotación económicamente ineficiente y pérdida de recursos naturales como suelo y agua.
- A pesar de existir técnicas accesibles para realizar el análisis del estado de los suelos estos requieren de conocimientos técnicos para su realización o son poco conocidas, lo cual ocasiona que los productores agrícolas no las utilicen.
- El análisis de los suelos realizado en laboratorios especializados conlleva un gasto significativo de recursos económicos y de tiempo ya que no existen laboratorios cercanos a las poblaciones rurales lo cual dificulta el seguimiento de la evolución de los lotes.
- Si bien los análisis químicos realizados en laboratorios proporcionan un diagnóstico cuantitativo de los componentes del suelo, no indican si estos se encuentran biológicamente disponibles para realizar una siembra, lo cual no determina un análisis completo de estos.

1.4 DEFINICIÓN DE OBJETIVOS

En base a la problemática planteada se definen el objetivo general y objetivos específicos en propuesta a la solución del problema planteado.

1.4.1 OBJETIVO GENERAL

Plantear un algoritmo de segmentación de imágenes de cromatografía de suelos para apoyar en el diagnóstico del estado de los mismos.

1.4.2 OBJETIVOS ESPECÍFICOS

- Evitar el uso inapropiado de los suelos
- Promover la práctica del análisis de los suelos para determinar su mejor tipo de uso.
- Facilitar el proceso de análisis de suelos para los agricultores.
- Reducir los costos del proceso del análisis de los suelos y el tiempo de obtención de sus resultados.
- Desarrollar un prototipo de aplicación móvil que ayude al diagnóstico del estado de los suelos

1.5 HIPÓTESIS

Hi: El uso del algoritmo de agrupamiento k-means para la segmentación de imágenes de cromatografía de suelos permite realizar un diagnóstico del estado de los mismos con una confiabilidad de por lo menos 80%.

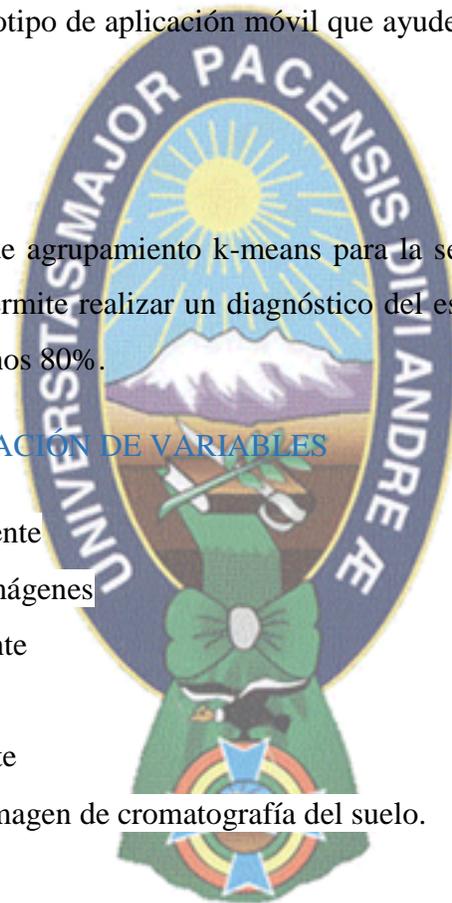
1.6 OPERACIONALIZACIÓN DE VARIABLES

- a) Variable independiente
 - Segmentación de imágenes
- b) Variable Interviniente
 - Algoritmo k-means
- c) Variable dependiente
 - Diagnóstico de la imagen de cromatografía del suelo.

1.7 JUSTIFICACIÓN

1.7.1 JUSTIFICACIÓN ECONÓMICA

Contar con una algoritmo que logre una buena calidad en la segmentación de imágenes de cromatografía de suelos que pueda ser implementado en distintas plataformas para apoyar en el diagnóstico cualitativo del estado de los suelos reducirá de manera considerable los



gastos de traslado de las muestras desde poblaciones rurales hacia laboratorios o de la contratación de personal especializado para realiza esta tarea lo que promoverá la práctica del análisis previo de los suelos antes de su uso y así evitar el desgaste de estos y así mejorar su producción.

1.7.2 JUSTIFICACIÓN SOCIAL

Esta investigación contribuirá con instituciones como la FAO, ECOTOP entre otras, en su búsqueda de concientización del uso sustentable de los suelos a través de la práctica del análisis de los mismos y promover el uso de la técnica de Cromatografía para concientizar a los pobladores de las áreas rurales acerca del daño que están sufriendo los suelos en todo el mundo por el uso indiscriminado de fertilizantes químicos.

También contribuirá a los agricultores de zonas rurales brindándoles una herramienta de uso fácil y rápido para realizar el análisis del estado de sus suelos antes de realizar cualquier tipo de uso de los mismos.

1.7.3 JUSTIFICACIÓN CIENTÍFICA

La propuesta de un algoritmo que contribuya al diagnóstico del estado de los suelos mediante la segmentación de imágenes de cromatografía abre una nueva línea de investigación interdisciplinaria que puede ser aprovechada tanto por profesionales del área de informática, como por profesionales del área de agronomía, Además que en su desarrollo se hace uso de conocimientos en las áreas de la programación gráfica, la matemática, manejo de estructuras de datos, almacenamiento, su estudio, desarrollo e implementación amplían y fortalecen los conocimientos sobre estas y otras áreas relacionadas.

1.8 ALCANCES Y LÍMITES

Para el logro de los objetivos planteados el presente trabajo se delimita bajo los siguientes alcances y límites:

1.8.1 ALCANCES

1.8.1.1 ALCANCE TEMPORAL

El presente Trabajo se desarrollará durante el segundo semestre de la gestión 2016.

1.8.1.2 ALCANCE ESPACIAL

Se eligió como área piloto para la obtención de imágenes digitales a la Estación Experimental de la Facultad de Agronomía y Zootecnia de la UMSA ubicado en la población de Sapecho (Alto Beni, La Paz).

1.8.2 LÍMITES

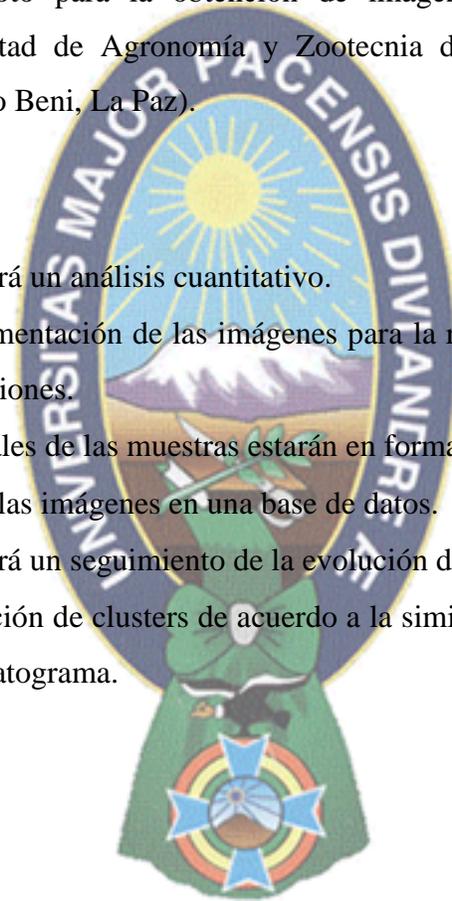
- La aplicación no hará un análisis cuantitativo.
- La salida de la segmentación de las imágenes para la representación de las mismas será tomada por regiones.
- Las imágenes digitales de las muestras estarán en formato BPM.
- No se almacenarán las imágenes en una base de datos.
- La aplicación no hará un seguimiento de la evolución de los suelos.
- Se hará la clasificación de clusters de acuerdo a la similitud de colores que presente la imagen del cromatograma.

1.9 APORTES

1.9.1 PRÁCTICO

El principal aporte práctico de la presente investigación es la implementación del algoritmo de agrupamiento k-mean para la segmentación de imágenes de cromatografía de suelos.

Otro aporte es el desarrollo de la aplicación móvil para el diagnóstico del estado de los suelos, apoyando así el proceso de análisis de suelos.



1.9.2 TEÓRICO

El principal aporte teórico del presente trabajo es contribuir a la investigación de la aplicación de Segmentación de imágenes de color en áreas como la agricultura mediante técnicas de *clustering* como el algoritmo *k-means*.

1.10 METODOLOGÍA

La presente investigación será de tipo descriptiva utilizando el método de investigación científica hipotético-deductiva siguiendo los lineamientos de Mario Bunge, aplicando la técnica de Procesamiento de imágenes, planteada por Gonzales y Woods (2002) siguiendo las siguientes etapas:

- Adquisición de Imágenes
- Pre-Procesamiento
- Segmentación
- Representación
- Reconocimiento

A su vez la etapa de segmentación será desarrollada en base a técnicas de *Clustering* no supervisadas más específicamente el Algoritmo *K-means*.

El prototipo de la aplicación móvil que permita evaluar la hipótesis planteada; se desarrollara en el entorno de desarrollo Android Studio 2.2.0



MARCO TEÓRICO

2.1 IMAGEN DIGITAL

La imagen digital es cualquier imagen fija o en movimiento, que se capture en un medio electrónico y que se represente como un archivo de información leído como una serie de pulsos eléctricos.

2.1.1 CONCEPTOS BÁSICOS DE IMAGEN DIGITAL

2.1.1.1 PIXEL

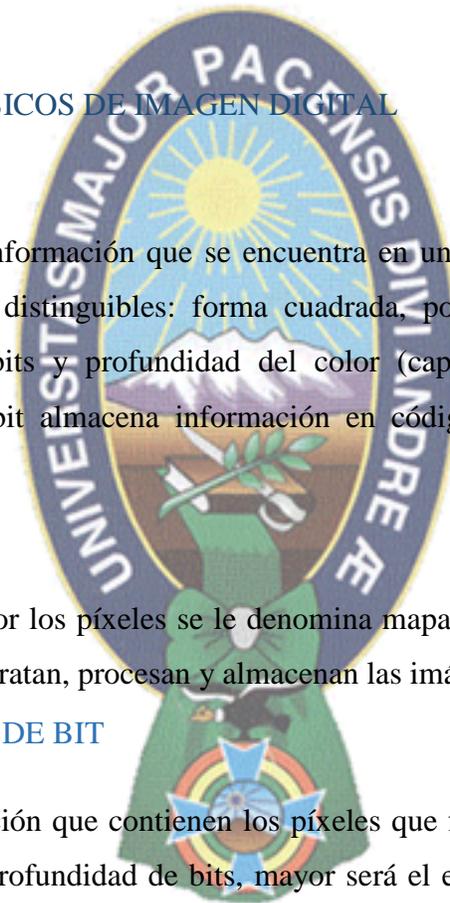
Es la unidad mínima de información que se encuentra en una imagen digital, los píxeles tienen tres características distinguibles: forma cuadrada, posición relativa a los demás píxeles en un mapa de bits y profundidad de color (capacidad de almacenar color) expresada en bits. Cada bit almacena información en código binario sobre su color y luminosidad.

2.1.1.2 MAPA DE BITS

A la cuadrícula formada por los píxeles se le denomina mapa de bits de la imagen digital. En este formato se editan, tratan, procesan y almacenan las imágenes digitales.

2.1.1.3 PROFUNDIDAD DE BIT

Es la cantidad de información que contienen los píxeles que forman el mapa de bits de la imagen digital. A mayor profundidad de bits, mayor será el espacio de almacenamiento y menor la velocidad de tratamiento de la imagen.



2.1.2 REPRESENTACIÓN DE IMÁGENES

Las imágenes según Alvarado (2012) pueden ser representadas en diferentes estructuras matemáticas como funciones, matrices, conjuntos y grafos, a continuación, se describe cada una.

2.1.2.1 IMÁGENES COMO FUNCIONES

Como señales, la primera representación utilizada para imágenes corresponde a funciones matemáticas. Las imágenes en espacio discreto se den en un dominio de coordenadas espaciales discretas y finitas, en forma de una rejilla rectangular. Dicho dominio es mapeado hacia un conjunto de valores vectoriales en \mathbb{R}^n , donde n equivale a la dimensión de la señal; por ejemplo, para imágenes monocromáticas $n = 1$, para imágenes convencionales a color $n=3$. Así, la función f definida como:

$$f: \mathbb{X} \rightarrow \mathbb{R}^n$$

representará una imagen en espacio discreto, donde $\mathbb{X} = \{0, 1, \dots, D - 1\}^d \subset \mathbb{N}^d$ es el conjunto de posiciones validas de los píxeles en un espacio de d dimensiones. Así $f(\underline{x})$ es una función que retorna el vector que representan la composición espectral de una imagen sobre la posición \underline{x} .

Si el codominio \mathbb{R}^n es remplazado por un conjunto discreto de valores \mathbb{V}^n entonces se tiene una imagen digital.

Obsérvese que en general las funciones que representan una imagen tienen carácter vectorial. Como canal k se denota una única componente $f_k(\underline{x})$, $k = 1, \dots, n$ de la imagen $f(\underline{x})$.

En el caso de imágenes bidimensionales es común implicar la notación del vector $\underline{x} = [x, y]^T$ como par ordenado (x, y) .

2.1.2.2 IMÁGENES COMO MATRICES

Para el caso particular del espacio bidimensional ($d = 2$), la forma de rejilla que asume el dominio de definición de las funciones utilizadas para representar imágenes y canales permite derivar una segunda forma de representación basada en matrices, particularmente apta para el manejo computacional de estas señales.

La matriz F representa a la misma señal dada por la función f si se cumple:

$$f_{ji} = f \left(\begin{bmatrix} i \\ j \end{bmatrix} \right)$$

Donde, además

$$F = \begin{bmatrix} f_{0,0} & f_{0,1} & \dots & f_{0,C-1} \\ f_{1,0} & f_{1,1} & \dots & f_{1,C-1} \\ \vdots & \vdots & \ddots & \vdots \\ f_{R-1,0} & f_{R-1,1} & \dots & f_{R-1,C-1} \end{bmatrix}$$

Con R igual al número de filas y C igual al número de columnas.

De nuevo si $f_{i,j} \in \mathbb{R}^n$ se tiene una imagen en espacio discreto, y si $f_{i,j} \in \mathbb{N}^n$ se dice que F es una imagen digital. Lo anterior se extiende de forma directa para canales.

2.1.2.3 IMÁGENES COMO CONJUNTOS

La diversidad de técnicas existentes para el procesamiento y análisis de imágenes han forzado a crear otros tipos de representaciones adicionales a los dos enunciados anteriormente.

Una representación frecuentemente usada es considerar a las imágenes como conjuntos de pixeles. De forma general un pixel p es una tupla $\langle \underline{x}, \underline{c} \rangle$ compuesta por un vector de posición \underline{x} y un vector \underline{c} que representa la posición espectral de la energía incidente en la posición del pixel. Es necesario contar con dos operadores que permitan extraer la posición

o el valor de un pixel. Estos operadores se denominarán operador de posición “pos” y operador de valor “val”, de modo que:

$$\text{pos}(p) = \text{pos}(\langle \underline{x}, \underline{c} \rangle) = \underline{x}$$

$$\text{val}(p) = \text{val}(\langle \underline{x}, \underline{c} \rangle) = \underline{c}$$

la imagen \mathfrak{I} está definida como:

$$\mathfrak{I} = \{p | p = \langle x, c \rangle, x \in \mathbb{X}, c = f(x)\}$$

Nótese que, con esta definición, no todos los valores posibles de posiciones \underline{x} deben estar necesariamente incluidos en la imagen. De este modo, aparecen operaciones que agregan o eliminan pixeles a la imagen, lo que no es directamente representable con funciones o matrices.

2.1.2.4 IMÁGENES COMO GRAFOS

Las tareas del análisis de imágenes y de visión por computador usualmente tienen como fin detectar estructuras u objetos visuales con determinadas características en las imágenes.

Para los procesos de detección de dichos objetos es usual encontrar estructuras de datos basadas en grafos, que se componen de nodos y aristas. Como nodos se utilizan los pixeles, definidos como tuplas de posición y valor espectral $\langle \underline{x}, \underline{c} \rangle$. Las aristas y sus pesos definen relaciones adicionales entre los pixeles, que pueden basarse tanto en comparaciones de sus posiciones \underline{x} como de sus valores \underline{c} .

2.1.3 HISTOGRAMA

Supongamos dada una imagen en niveles de grises, siendo el rango de 256 tonos de gris (de 0 a 255). El histograma de la imagen consiste en una gráfica donde se muestra el número de píxeles, n_k , de cada nivel de gris, r_k , que aparecen en la imagen.

El análisis estadístico derivado del histograma puede servir para comparar contrastes e intensidades entre imágenes. El histograma podría ser alterado para producir cambios en la imagen. Por ejemplo, el histograma es utilizado para binarizar una imagen digital, es decir, convertirla en una imagen en blanco y negro, de tal manera que se preserven las propiedades "esenciales" de los Histogramas imagen. La forma usual de binarizar una imagen es eligiendo un valor adecuado o umbral, u , dentro de los niveles de grises, tal que el histograma forme un "valle" en ese nivel. Todos los niveles de grises menores que u se convierten en 0 (negro), y los mayores que u se convierten en 255 (blanco).

2.2 ÁREAS RELACIONADAS CON IMÁGENES

En los últimos años el uso del computador para el tratamiento imágenes se ha especializado y de ahí que se pueden distinguir seis áreas (Alvarado, 2012):

- Adquisición de imágenes
- Procesamiento digital de imágenes
- Análisis digital de imágenes
- Comprensión de imágenes
- Visión por computador
- Visión industrial

Aunque en general entre la mayor parte de los autores en el área aún no existe un consenso de en donde inicia y en donde termina cada una de estas áreas es común colocar en un extremo al Procesamiento de imágenes y en el otro a la visión por computador. En la figura 2.1 se puede observar estas áreas, mostrando la diferencia que existe entre ellas de acuerdo a la información de entrada y la salida.

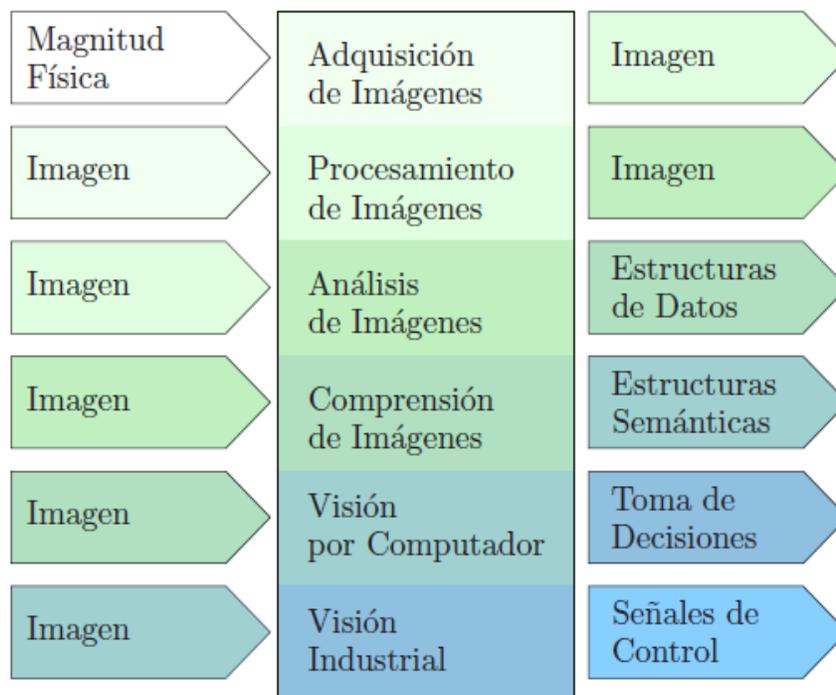


Figura 2.1 Gama de áreas relacionadas con el tratamiento de imágenes

Fuente: Alvarado, 2012

Para el caso particular de esta investigación nos centraremos en la especialidad de Procesamiento digital de Imágenes.

2.3 PROCESAMIENTO DIGITAL DE IMÁGENES

El procesamiento digital de imágenes (PDI) consiste en la manipulación de los datos contenidos en la imagen para convertirlos en información útil. Dicho procesamiento puede realizarse en forma óptica, analógica o digital. El procesamiento de imágenes mediante medios de procesamiento digital de la información, tales como PC, FPGA (*Field Programmable Gate Array*), aplicaciones específicas sobre DSP (*Digital Signals Processor*), etc.; constituye lo que se denomina Procesamiento Digital de Imágenes (PDI) (Gonzalez, y Woods, 2002).

El PDI se realiza sobre una representación digital de la imagen. Esa representación está basada en la cuantificación y discretización de la información de color contenida y en la división espacial de la imagen en diferentes regiones. Estas regiones están representadas por píxeles (Comas y Meschino, 2014).

Las técnicas de PDI pueden definirse como cualquier técnica cuya entrada sea una imagen y la salida puede ser, ya sea una imagen o las características y parámetros relacionados a la imagen de entrada, entonces el PDI puede analizarse desde dos enfoques diferentes: desde el punto de vista del análisis del problema llevado a cabo o desde el punto de vista de la etapa del procesamiento (Gonzales y Woods 2002; Comas y Meschino 2014).

Entonces desde el punto de vista del análisis del problema el PDI puede ser realizado de acuerdo a tres líneas del pensamiento: heurístico que utiliza herramientas de la inteligencia computacional, determinístico que realiza un análisis cuantitativo empleando herramientas algebraicas o morfológicas y probabilístico que utiliza modelos estadísticos ya sea de reconocimiento de patrones o redes neuronales (Comas y Meschino, 2014).

<u>Etapas del PDI</u>		<u>Nivel de abstracción</u>
	Pre-Procesamiento	<i>Bajo</i>
	Segmentación	<i>Medio</i>
Análisis {	Representación	
	Interpretación	

Figura 2.2. Etapas del Procesamiento Digital de Imágenes
Fuente: Comas y Meschino, 2014

Desde el punto de vista de la etapa de resolución del problema se pueden distinguir cuatro etapas esenciales (Gonzales y Woods, 2002 en Comas y Meschino, 2014): el pre-procesamiento, segmentación, representación e interpretación. Estos a su vez se clasifican según su nivel de abstracción en procesos de bajo nivel, de nivel medio y de alto nivel (ver figura 2.2).

2.3.1 PROCESAMIENTO DE BAJO NIVEL

A este nivel de abstracción pertenecen la etapa de adquisición y el pre-procesamiento de imágenes. Este nivel de abstracción se caracteriza por que sus entradas son imágenes y sus salidas también.

2.3.1.1 ADQUISICIÓN DE LA IMAGEN

Como su nombre lo indica en esta etapa se adquiere la representación pictórica de un objeto en medio digital y es el primer paso para el PDI.

2.3.1.2 PRE-PROCESAMIENTO

En esta etapa se realizan procedimientos en la imagen para reducir el ruido, mejorar el contraste y filtros de enfoque. Por otra parte, en esta etapa también se destacan las áreas de interés de la imagen, se mejora la apariencia de la imagen.

2.3.2 PROCESAMIENTO DE NIVEL MEDIO

A este nivel de abstracción pertenece la etapa de Segmentación y la de Representación. Esta se caracteriza por que sus entradas por lo general son imágenes, pero sus salidas son atributos que han sido extraídos de esas imágenes (contornos, bordes, patrones, similitud de colores, entre otros).

2.3.2.1 SEGMENTACIÓN

Se refiere a los procedimientos que particionan una imagen en sus partes constituyentes u objetos. Esta es la tarea más difícil del PDI ya que es donde se implementan los algoritmos que pueden o no llevar a un resultado satisfactorio; en general mientras más precisa sea la

segmentación es más probable que el reconocimiento tenga éxito. De esta tarea generalmente se obtienen son datos de pixeles en bruto. En Particular para esta etapa se utilizarán Técnicas de *Clustering* más precisamente el algoritmo *k-means*.

2.3.2.2 REPRESENTACIÓN

En esta tarea la primera decisión es si los datos de salida de la segmentación deben ser tomados como límites o como una completa región. La representación como límite es apropiada cuando la atención se centra en las características de forma externa, tales como esquinas y las inflexiones. La representación como región es adecuada cuando la atención se centra en las propiedades internas, tales como textura o la forma. En algunas aplicaciones, estas representaciones se complementan entre sí. La descripción, también llamada función de selección, se refiere a la extracción de atributos que resultan en alguna información cuantitativa de interés o son básicos para la diferenciación de una clase de objetos de otro.

2.3.3 PROCESAMIENTO DE ALTO NIVEL

A este nivel de abstracción pertenece la etapa de reconocimiento, la cual implica el obtener algún significado de un conjunto de objetos reconocidos y, finalmente, realizar las funciones cognitivas asociadas con la vista.

2.3.3.1 RECONOCIMIENTO

Es el proceso de darle una etiqueta a la imagen por ejemplo “vehículo”. Este proceso se basa en el conocimiento acerca de un dominio del problema; este conocimiento se codifica en un sistema de procesamiento de imágenes en forma de una base de datos de conocimiento. Este conocimiento puede ser tan simple como que detalla las regiones de una imagen donde se conoce la información de interés que se encuentra, por lo tanto, limitar la búsqueda que tiene que ser llevado a cabo en la búsqueda de esa base de conocimientos información. Este también puede ser bastante complejo, tal como una lista interrelacionada de los principales defectos posibles en un problema inspección de materiales o de un banco

de imágenes que contienen imágenes de satélite de alta resolución de una región en relación con aplicaciones de detección de cambios. Además de guiar la operación de cada módulo de procesamiento, la base de conocimientos también controla la interacción entre los módulos.

2.4 SEGMENTACIÓN DE IMÁGENES

La segmentación de imágenes según Gonzales y Woods (2002) puede definirse como la agrupación de píxeles de distintas intensidades de colores similares en un espacio paramétrico.

Para definir formalmente la tarea de segmentación se requiere retomar la notación de imágenes como conjuntos como se detalló anteriormente, así como las siguientes definiciones.

Una región \mathcal{R} en una imagen es un subconjunto no vacío de la imagen \mathcal{I} , esto es $\mathcal{R} \subseteq \mathcal{I}, \mathcal{R} \neq \emptyset$. Una región no requiere ser topológicamente conexa, y por tanto dos partes visiblemente separadas pertenecientes a un objeto parcialmente oculto detrás de otro pueden pertenecer a la misma región.

La partición \mathcal{P} de una imagen \mathcal{I} es un conjunto de n regiones $\{\mathcal{R}_i; i = 1 \dots n\}$ tales que $\cup_{i=1}^n \mathcal{R}_i = \mathcal{I}$ y $\mathcal{R}_i \cap \mathcal{R}_j = \emptyset$ para $i \neq j$. Una partición se dice ser más fina que otro si parte a la imagen en regiones más pequeñas.

A nivel de imágenes no existe concepto para el término objeto. En vez de ello, se determinan componentes o partes utilizando un predicado de uniformidad u homogeneidad H_i para alguna característica específica, y otro predicado que evalúa la adyacencia de regiones A .

La segmentación basada en imagen S_i de una imagen \mathcal{I} es una partición de \mathcal{I} que satisface para cada región $\mathcal{R}_i \in S_i, H_i(\mathcal{R}_i)$, y $\neg H_i(\mathcal{R}_i \cup \mathcal{R}_j)$ para $A(\mathcal{R}_i, \mathcal{R}_j)$.

La condición $H(\mathcal{R}_i \cup \mathcal{R}_j)$ previene partir la imagen en demasiadas regiones (sobre segmentación): si el resultado de fusionar cualquier par de regiones adyacentes es inhomogeneo, entonces eso implica que la segmentación ya contiene el mayor número de regiones que satisfacen el predicado de homogeneidad. Todas las estrategias de segmentación a nivel de imagen se pueden considerar como una implementación particular del predicado H. En ellos se agregan restricciones adicionales a la definición región, permitiendo especificar aún más las propiedades que debe cumplir una región valida. Por ejemplo, una condición usualmente impuesta a H es la conectividad de las regiones.

La forma de realizar la segmentación depende del tipo de imagen particular y de la información que se tiene de la misma. Así, por ejemplo, puede distinguirse entre: segmentación manual (por parte de un experto), segmentación basada en técnicas de procesamiento de imágenes, segmentación través de técnicas de reconocimiento de patrones o mediante inteligencia computacional (Comas y Meschino, 2014). Para el caso particular de este documento nos centraremos en la Segmentación a través de reconocimiento de patrones.

2.5 TÉCNICAS DE RECONOCIMIENTO DE PATRONES

La segmentación mediante reconocimiento de patrones comprende dos etapas: en la primera se realiza un procesamiento de las imágenes mediante el cual se extraen diferentes características (*features*) para cada píxel. Éstas pueden ser características de texturas, información de color, información de posición, entre otras. (Meschino et al., 2008; Meschino et al., 2006b; Moler et al., 1999). En la segunda etapa, las características son ingresadas a un algoritmo de reconocimiento de patrones. Por medio de este algoritmo, se obtiene el conjunto de datos de entrada, que representa a los píxeles de la imagen original, agrupado o clasificado en diferentes conjuntos o etiquetas, lo que permitirá generar una imagen segmentada. La selección de las características o descriptores a utilizar es una tarea sumamente importante del proceso, la cual debe realizarse de acuerdo a la problemática planteada y de ella dependerá en gran medida el desempeño del proceso de segmentación.

El reconocimiento de patrones se define como el proceso en el cual se agrupa un conjunto de datos en una determinada cantidad de grupos (clusters) o etiquetas mediante un mecanismo de decisión (Jain et al., 1999). Este mecanismo depende de la técnica particular que se utilice y es el que define al algoritmo de reconocimiento de patrones. De acuerdo a la información que se utilice para el ajuste o definición de la técnica de reconocimiento de patrones, puede distinguirse entre (Haykin, 1999; Jain et al., 1999):

- Técnicas supervisadas: En ellas los parámetros internos del algoritmo son ajustados de acuerdo a un conjunto de datos previamente clasificado (un conjunto de datos etiquetados en diferentes clases). La descripción obtenida de este proceso de ajuste es, entonces, utilizada para clasificar un nuevo conjunto de patrones. Este tipo de técnicas se conoce como clasificación supervisada o análisis discriminante.
- Técnicas no supervisadas (clustering): En este tipo de técnica, el problema consiste en agrupar un conjunto de datos en grupos significativos, sin utilizar información previa sobre el agrupamiento esperado. La información de las clases es únicamente obtenida a partir de los propios patrones a agrupar. Este tipo de técnicas se conocen como técnicas de clasificación no supervisada, agrupamiento o clustering.

2.6 ALGORITMO K-MEANS

El algoritmo k-means es una técnica de clustering para la segmentación de imágenes digitales. En este algoritmo se define el número de clusters k a generar. Por cada *cluster*, se tendrá un centro en el espacio de patrones. Los datos son asignados a uno de los k clusters según su distancia al centro sea la menor. El algoritmo se basa en la minimización de la suma de los cuadrados de las distancias de todos los puntos de un *cluster* al centro del mismo, es decir, en la minimización de la función criterio de errores cuadráticos. El algoritmo puede definirse como sigue (Jain et al., 1999):

1. Se toman k centros de *cluster* iniciales $\{Z_j\}_{j=1,\dots,k}$. El criterio para asignar los centros iniciales es arbitrario.

2. En la j -ésima iteración se distribuyen el conjunto de patrones X entre los k centros, asignando cada punto al cluster cuyo centro esté más cercano. Para esta asignación pueden utilizarse diferentes definiciones de distancia (distancia euclidiana, suma del valor absoluto de las diferencias, etc.).

3. Se calculan los nuevos centros de cluster siguiendo el criterio de error cuadrático. Es decir, se busca minimizar el índice E dado por:

$$E_j = \sum_{X \in S_{j,i}} \|X - Z_i(j+1)\|^2$$

donde E es el error cuadrático; $X \in S_{j,i}$ es conjunto de patrones asignados al cluster i en la iteración j y Z_i es el centro del cluster i .

Los nuevos centros que minimizan el índice dado por 3.1 resulta ser el medio entre los patrones asignados a cada cluster, y estarán dados por:

$$Z_i(j+1) = \frac{1}{N_j} \sum_{K \in S_i} X_k$$

donde i es el índice del cluster; $k \in S_i$ representa al conjunto de patrones asignados al clúster i ; N_i es el número de patrones en el cluster i y j es número de iteración.

4. Si $Z_i(j+1) = Z_i(j)$ para todos los clusters, el algoritmo se detiene. En otro caso se vuelve al paso 2). Existen metodologías que permiten obtener automáticamente la cantidad de clusters óptima para un determinado problema (Fraley and Raftery, 1998; Zhao et al., 2008). Se basan en ejecutar varias veces el algoritmo de clustering variando la cantidad de clusters k . Luego, por medio de una medida que permite evaluar la adecuación del agrupamiento obtenido con respecto a un modelo de distribución de datos predefinido, la cantidad y el agrupamiento óptimo es definido para el problema bajo estudio.

Una mejora realizada a este algoritmo planteada por Toscano, Pereira y Arzola (2010), consiste en calcular inicialmente el histograma de la imagen y hallar la frecuencia de

ocurrencia por cada color f_i . Luego en la ecuación de actualización de los centroides cada color es multiplicado por su frecuencia de ocurrencia en la imagen original, a fin de garantizar que la contribución de los píxeles con este color ejerza una influencia en el centroide de acuerdo con su frecuencia de aparición en la imagen.

2.7 CROMATOGRAFÍA DE SUELOS

Los inicios de la cromatografía datan de principios del siglo XX, en 1903 el botánico ruso Mijaíl Tswett, separó los pigmentos vegetales que dan color a las hojas. Sus disoluciones se hacían pasar por una columna de vidrio rellena de carbonato de calcio o cal finamente dividida, la cual interacciona de manera diferente con los componentes de la mezcla y, por tanto, estos se separan en distintas bandas coloreadas a lo largo de la columna. Para los años 20 el bioquímico Ehrenfried Pfeiffer de origen alemán, se dedicó al estudio de la calidad de los suelos y los alimentos de la agricultura industrial, quien observó que un suelo fértil contiene un conjunto complicado de microorganismos que continuamente crean, transforman y descomponen complejas moléculas orgánicas, a raíz de estos estudios desarrollo la técnica de Cromatografía de suelos (Restrepo y Pinheiro,2011).

La cromatografía en papel es una técnica analítica de carácter cualitativo, que se utiliza para dividir los distintos componentes de un producto. Aplicándolo a los estudios de Pfeiffer, este método consiste en recrear el suelo o el abono y lo que sucede en él a través de una imagen (“cromos” significa imagen, y “grafía” escritura) revelada en un papel filtro impregnado con nitrato de plata al 0,5%. Dicha imagen nos ofrece gran cantidad de información sobre el estado de salud y la calidad del suelo o del abono, información cualitativa sobre la presencia de microorganismos, minerales y materia orgánica, así como su grado de actividad e interacción entre ellos. Es decir, que no se puede decir mediante una cromatografía que en un suelo hay un 0,8% de materia orgánica (análisis cuantitativo), sin embargo, se puede saber si hay mucha o poca materia orgánica, si ésta es de calidad, si está disponible para los microorganismos para que éstos la descompongan, si los microorganismos tienen acceso. (Restrepo y Pinheiro, 2011; Torres, 2013).

El cromatograma recoge toda la información vital del suelo, dónde se puede leer la interacción de los factores biológicos, químicos y físicos entre ellos y con el medio. Esta interacción nos muestra el grado de calidad que posee el suelo, y si cada uno de los factores está en armonía con los demás (Restrepo y Pinheiro,2011).

En base a estos tres factores, la calidad del suelo se compone de tres elementos esenciales para un adecuado funcionamiento (las 3M): los microorganismos, los minerales y la materia orgánica. El complejo suelo que posea una armonía entre estos tres elementos, será un suelo de máxima calidad. (Restrepo y Pinheiro,2011).

2.8 INTERPRETACIÓN DE CROMATOGRAMAS

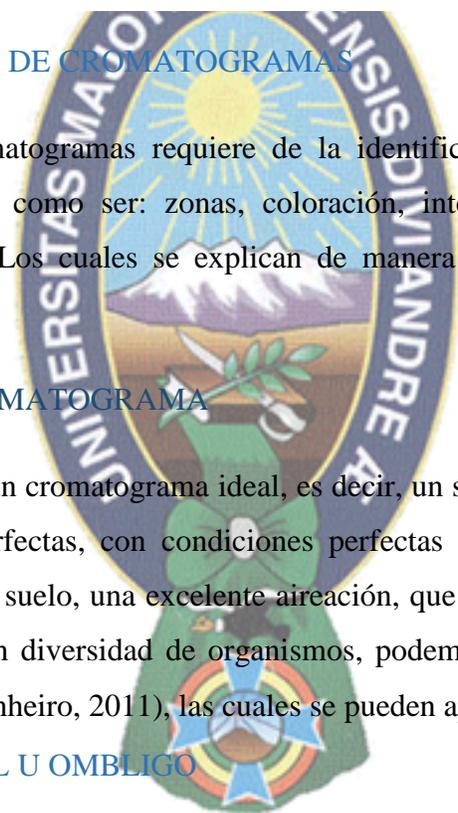
La interpretación de cromatogramas requiere de la identificación de ciertos elementos presentes en los mismos, como ser: zonas, coloración, integración de zonas, formas, patrones. (Torres, 2013). Los cuales se explican de manera detallada en los siguientes puntos.

2.8.1 ZONAS DEL CROMATOGRAMA

Suponiendo la imagen de un cromatograma ideal, es decir, un suelo que se encuentre en las condiciones climáticas perfectas, con condiciones perfectas de humedad, una excelente estructura microporosa del suelo, una excelente aireación, que cuente con aporte constante de materia orgánica y gran diversidad de organismos, podemos distinguir claramente las cinco zonas (Restrepo y Pinheiro, 2011), las cuales se pueden apreciar en la figura 2.3.

2.8.1.1 ZONA CENTRAL U OMBLIGO

En esta zona está formada por todo el lavado al que ha sido sometida esta zona, ya que toda la solución que se extiende por el papel filtro pasa por ahí y el nitrato de plata se ha combinado en su totalidad. (Torres, 2013; Restrepo y Pinheiro, 2011).



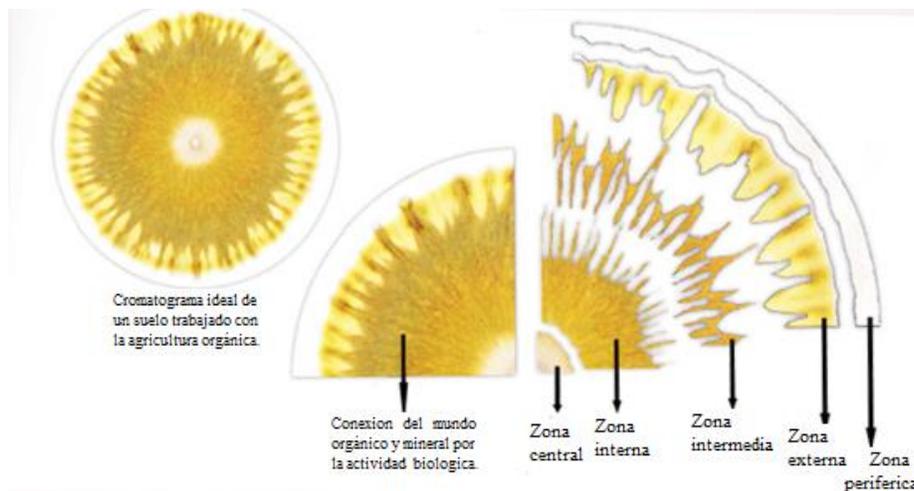


Figura 2.3. Identificación de las zonas que forman un cromatograma ideal
Fuente: Restrepo y Pinheiro, 2011

2.8.1.2 ZONA INTERNA O MINERAL

Los complejos arcillosos, alófanos, y arcillo-húmicos se desplazan más lentamente que los complejos de sólo materia orgánica particulada, y éstos a su vez más lentos que el humus (sensu stricto).

Como hay abundantes y variados complejos orgánicos, en esta zona quedan los unidos a arcillas y alófanos, dando una textura grumosa. El agua de la solución en su movimiento centrífugo, desprovista de estos complejos por la adhesión de unos con otros, forma una red de canalillos que asemejan la estructura de una pluma, con bábulas que confluyen en un raquis. (Restrepo y Pinheiro, 2011).

2.8.1.3 ZONA INTERMEDIA U ORGÁNICA

Aquí se depositan complejos orgánicos sin vinculación con materia mineral, y por tanto menos densos y más fácilmente transportables, quedando depositados más externamente y

formando esta zona, dejando también una textura grumosa. Los canalillos anteriormente dichos se continúan en esta zona. (Torres,2013; Restrepo y Pinheiro, 2011).

2.8.1.4 ZONA EXTERNA O HÚMICA

Finalmente quedan los agregados más livianos que son los componentes del humus en sentido estricto. El raquis de algunas plumas llega hasta el final de la zona. El humus permanente forma manchas o lunares. (Torres,2013; Restrepo y Pinheiro, 2011).

2.8.1.5 ZONA PERIFÉRICA O SECA

Es la zona del papel a la que no ha llegado la solución. Permite ver el cromatograma. Se utiliza para realizar anotaciones que identifiquen el cromatograma y para manipularlo evitando interferencias en la imagen que se forma. (Torres,2013; Restrepo y Pinheiro, 2011).

2.8.2 COLORACIÓN DEL CROMATOGRAMA

Es fruto de las reacciones que se producen entre los componentes del suelo con el hidróxido de sodio y el nitrato de plata. También de la reacción entre el nitrato de plata con el hidróxido de sodio. La reacción del nitrato de plata con la materia orgánica desemboca en productos con plata. La luz produce una reducción de estos compuestos, precipitándose la plata y dando coloraciones en la gama de los ocre (Torres,2013).

La reacción del nitrato de plata con el hidróxido de sodio, produce óxido de plata de color negro. Cuando hay poca materia orgánica en el suelo, la solución que corre por el papel contiene hidróxido de sodio en más o menos cantidad dependiendo de la abundancia de materia orgánica, ya que, en la preparación de la solución, el hidróxido de sodio ha tenido oportunidad de combinarse con ella (Torres,2013; Restrepo y Pinheiro, 2011).

Los colores que reflejan el buen estado evolutivo y saludable tanto de los suelos como de los abonos son amarillo, dorado, anaranjado, rojizo o café claro y tonalidades verdosas. Cuando en un cromatograma encontramos la combinación de cafés claros y muy oscuros, estamos

ante un proceso intermedio de desarrollo, en el cual la materia orgánica está cruda, acumulada o en etapa de maduración e integración en el abono o al suelo. Por otro lado, los colores Negro, ceniza, pardo muy oscuro, lilas o violeta, gris y tonalidades azuladas reflejan un mal estado evolutivo y no saludable de los suelos y de los abonos orgánicos procesados o en proceso (Restrepo y Pinheiro, 2011)

2.8.3 INTEGRACIÓN DE ZONAS

Se aprecia en los límites interzonales. La integración es una cualidad que varía entre dos extremos: Límites netos y límites difusos, estando correlacionados respectivamente con el aislamiento o la conexión de las partículas. En el ecosistema clímax de suelo, es donde hay más conexión e integración entre los distintos componentes. Esto se muestra en el croma con una mayor difusión en los límites, especialmente entre zonas, mostrándose difuminados y de difícil discriminación. Un buen suelo mostrará gran integración entre las zonas, mientras que un suelo incipiente o en malas condiciones, no presentará zonas o si las presenta sus límites serán netos y no integrados. La presencia de anillos es una muestra de falta de integración entre las diversas partículas del suelo (Torres, 2013; Restrepo y Pinheiro, 2011).

2.8.4 FORMAS

Entendemos por formas a las figuras tomadas individualmente. Distinguimos ciertas figuras que llamaremos plumas, cabos – ensenadas, y manchas (Torres, 2013).

2.8.4.1 PLUMAS

Figuras semejantes a plumas de ave. Son originadas por el desplazamiento de la solución en el papel de filtro, liberada de agregados moleculares que van quedando atrapados en las inmediaciones de la corriente. Llamamos raquis, por semejanza con la pluma, a la estructura lineal radial central en la que confluyen las diminutas corrientes o bárbulas. Su ausencia en el cromatograma indica carencia de agregados en la solución.

Un buen suelo presentará plumas con bárbulas interconectadas, de límites difusos, y con raquis anchos, irregulares en su trayectoria y profundamente hundidos en la zona externa. Los raquis estrechos y rectilíneos y las bárbulas -cuando existen- son cortas y no interconectadas con las de las plumas adyacentes, son síntoma de un suelo en condiciones no deseables, bien por carencias en materia orgánica o por condiciones deficientes en la misma. (Restrepo y Pinheiro,2011).

2.8.4.2 CABOS Y ENSENADAS

Se muestran únicamente en el área externa del cromatograma, y los llamamos así por la semejanza con estos accidentes geográficos. Los cabos pertenecen a la zona intermedia. Las ensenadas conforman la zona externa. Los cabos son los extremos que pueden interrumpirse en la zona periférica. Corresponden con el final de la corriente de solución, y participan de las cualidades enunciadas anteriormente para los raquis. (Restrepo y Pinheiro,2011).

Los cabos anchos, largos, desiguales tanto en anchura como en longitud, indican buenas condiciones del suelo. Por el contrario, las formas aciculares, cortas, uniformes, nos indican condiciones deficientes. En las ensenadas se encuentran diversos componentes del humus.

2.8.4.3 MANCHAS

Nos referimos a estas figuras en el sentido habitual del término. Cuando aparecen lo hacen en la zona húmica. Se encuentran independientemente de los cabos o formando parte de ellos. Presentan coloración en el rango deseable, ocre, café oscuro, rojizo. Las manchas color café oscuro se presentan siempre en suelos fértiles. Podrían corresponder con el humus permanente. Su ausencia nos indica condiciones no deseables para el suelo (Restrepo y Pinheiro,2011).

2.8.5 PATRONES

Entendemos por patrón a la cualidad de un conjunto unitario de formas que hace referencia a las similitudes de los elementos o formas. Algunos patrones de interés para la interpretación de los cromatogramas serían los relativos a la uniformidad, los límites y la textura (Torres, 2013; Restrepo y Pinheiro, 2011).

Los patrones de uniformidad y límites, se definen en relación con su posición entre dos modos extremos. La textura hace referencia a la estructura fina de la porción de cromatograma que consideremos.

2.8.5.1 UNIFORMIDAD (REPETITIVO VS. VARIADO)

Se refiere a la igualdad de formas. Cuando las formas son todas iguales se alcanza la máxima uniformidad. Hablamos de un patrón repetitivo o uniforme. Corresponde con condiciones no deseables del suelo, que tendría muy poca diversidad, sobre todo biológica. Cuando todas las formas son distintas, nos hallaríamos con el máximo de variedad. Hablaríamos de un patrón variado. Corresponde con condiciones deseables en el suelo, de gran riqueza y diversidad de elementos que lo integran (Restrepo y Pinheiro, 2011).

2.8.5.2 LÍMITES (NETOS VS. DIFUSOS)

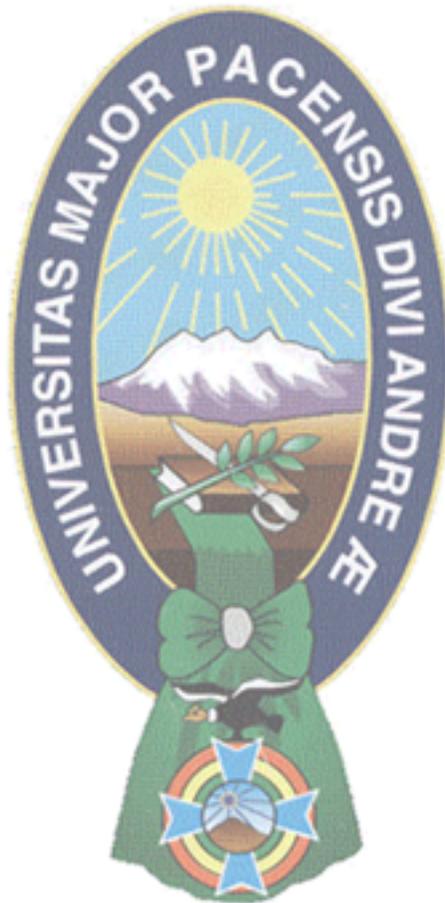
Se refiere a los límites de las formas, no a los límites de las zonas, si bien comparten con estos últimos la interpretación. Los límites netos son indicadores de separación y no integración en conjuntos mayores. Constituyen condiciones no deseables para los suelos si queremos tener condiciones climáticas de los mismos, mientras que los límites difusos indican lo contrario, siendo deseable esta condición para los suelos (Torres, 2013; Restrepo y Pinheiro, 2011).

2.8.5.3 TEXTURA (LISA VS. GRUMOSA)

Cualidad que pertenece a la conformación general de la sección del cromatograma que consideremos. Corresponde a la estructura fina o la estructura última que podemos

visualizar. Los extremos que consideramos son su igualdad (textura lisa) o desigualdad (textura grumosa).

Las texturas lisas nos indican condiciones abióticas predominantes en el suelo, transportándose en la solución moléculas inorgánicas pequeñas y no floculantes. Las texturas grumosas son síntoma de agregados orgánicos y floculación abundante. Esta textura es deseable (Torres,2013; Restrepo y Pinheiro, 2011).



MARCO APLICATIVO

3.1 INTRODUCCIÓN

Los expertos en manejo sostenible de suelos, están en una constante búsqueda de concientizar a los agricultores de realizar la práctica de análisis previo del estado de los suelos, con el objetivo de mejorar el estado de los mismos, una de las practicas que se busca inculcar, para este objetivo, es el análisis a través de la cromatografía en papel, por el hecho de ser un método económico, pero sin embargo de difícil comprensión. Por tal motivo la presente investigación plantea un algoritmo

En el presente capítulo se describe el procesamiento de imágenes de cromatografía de suelos siguiendo los lineamientos de la metodología de PDI planteada por Gonzales y Woods (2002), planteando el algoritmo de agrupamiento k-means, el cual nos permite realizar la segmentación de dichas imágenes, para lograr apoyar el diagnóstico del estado de los suelos.

Antes de llevar a cabo el proceso de segmentación, se realiza el pre-procesamiento de las imágenes para obtener de estas, otras con menos ruido, esto se logra a través de operaciones de realce, permitiendo una imagen de mejor calidad y más fácil de procesar en las siguientes etapas.

A partir de este tipo de imagen se realiza el proceso de segmentación a través del cual se particionan las imágenes, de tal manera que se pueden distinguir las cuatro zonas de interés de un cromatograma, zona central, interna, intermedia y externa, excluyendo previamente la zona periférica, ya que esta no llega a brindar información relevante del estado del suelo, a través de un algoritmo sencillo de umbralización. Esto pretende mejorar el tiempo de ejecución del algoritmo central. La figura 3.1 el diagrama de secuencia muestra el flujo del proceso.

Finalmente, tomando en cuenta la evaluación de la segmentación para cada valor de k , y habiendo evidenciado que la mejor segmentación se obtiene con $k=4$, se implementó el algoritmo k -means en un prototipo de una aplicación para dispositivos móviles, que realiza la segmentación de las imágenes de cromatografía de suelos capturadas con la cámara del dispositivo móvil, que permite determinar el diagnóstico del estado de los suelos, el cual se desarrolló en el IDE Android Studio.

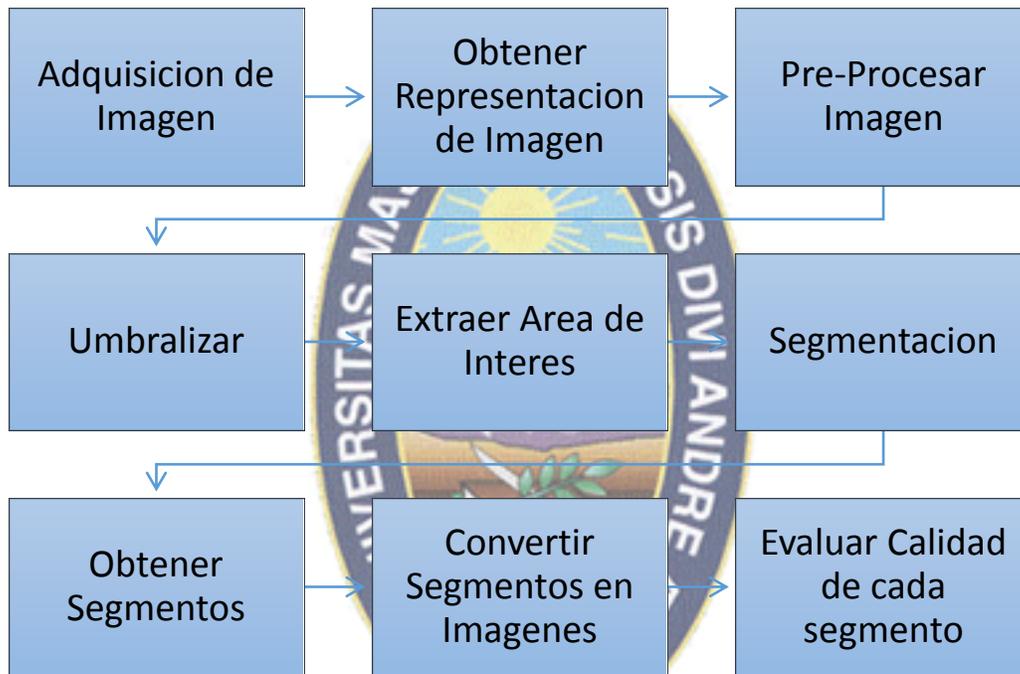


Figura 3.1 Diagrama de secuencia del procesamiento digital de imágenes de Cromatografía de suelos

Fuente: Elaboración propia

3.2 DESCRIPCIÓN DEL MÉTODO

EL método propuesto para el procesamiento de imágenes de cromatografía de suelos comprende las siguientes etapas:

- Adquisición de imágenes
- Pre-procesamiento
- Segmentación

- Representación
- Reconocimiento

Para las etapas críticas de esta investigación, es decir desde el Pre procesamiento hasta el reconocimiento, se realizan tareas que se eligieron según las necesidades particulares de esta investigación. En la figura 3.2. se muestra el diagrama de bloques de las tareas que se llevarán a cabo para cada etapa del método.

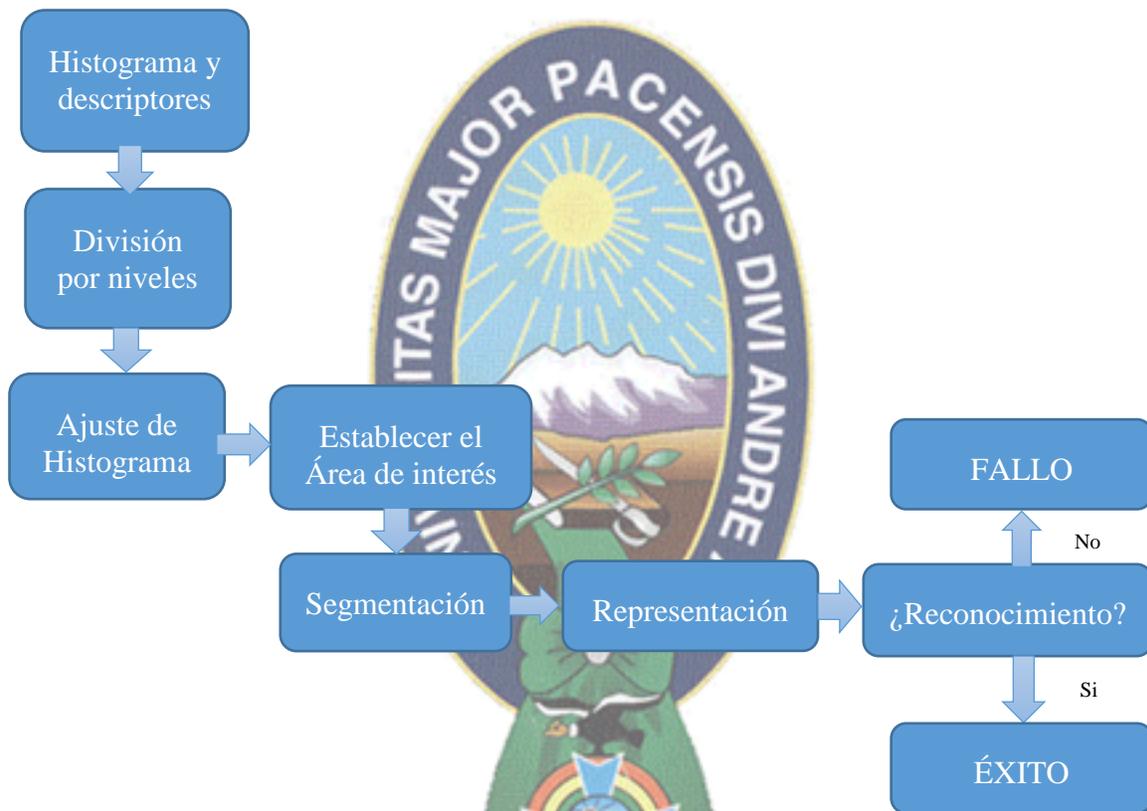


Figura 3.2 Diagrama de bloques del método propuesto
Fuente: Elaboración propia

3.2.1 ADQUISICIÓN DE IMÁGENES

En esta etapa se han determinado la formación, captura, digitalización, representación, y el almacenamiento de la imagen.

3.2.1.1 FORMACIÓN DE LA IMAGEN

Siendo la formación de la imagen la parte análoga de esta etapa, no se considera relevante para la investigación, sin embargo, cabe señalar que las imágenes de cromatografía de suelo se han formado a través de un proceso químico que resulta en una imagen similar al iris de un ojo plasmada en un papel filtro.

3.2.1.2 CAPTURA Y DIGITALIZACIÓN

Posterior a la formación las imágenes para ser procesadas serán capturadas y digitalizadas a través de la cámara digital de un dispositivo móvil de 8 megapíxeles.

3.2.1.3 REPRESENTACIÓN

La representación de las imágenes será como matriz bidimensional i con un tamaño de $640*480$, donde cada celda de la matriz representará un pixel, y $i(x,y)$ será el valor del pixel en la columna x y la fila y .

Tratándose en esta investigación con imágenes a color se determina:

$$1 \text{ pixel} = 3 \text{ Bytes}$$

Donde cada pixel consta de tres valores rojo, verde y azul (1 byte por color), lo que representa tres canales; es decir se trabaja en el modelo RGB con entrelazado de canales.

Se utilizarán el origen de las coordenadas de tipo *Bottom-left*, es decir el pixel $i(0,0)$ se encontrará en la esquina inferior izquierda, lo que quiere decir que trabajaremos en el segundo cuadrante del plano cartesiano.

3.2.1.4 ALMACENAMIENTO

El almacenamiento de la imagen será en formato BMP (*Bitmap*) ya que usa compresión sin pérdida RLE (*Run Length Encoding*) ya que para este caso no debe haber pérdida de calidad y además nos permite almacenamiento *Bottom-left* y entrelazado de canales.

3.2.2 PRE-PROCESAMIENTO

Como se mencionó en el capítulo anterior el objetivo de esta etapa del PDI es obtener una imagen con menos ruido para su fácil procesamiento en las etapas posteriores. Para el caso particular de esta investigación se llevarán a cabo tareas de pre-procesamiento como: calcular el histograma y descriptores de la imagen, calcular el umbral óptimo para una división por niveles y realizar un ajuste en el histograma para una división por canal.

3.2.3 HISTOGRAMA Y DESCRIPTORES

El cálculo del histograma representara una operación muy costosa desde el punto de vista computacional ya que se debe recorrer la imagen completa para determinar todos los valores de gris de la imagen y otros descriptores como la mediana y la desviación estándar.

Sin embargo, como se menciona en el capítulo anterior, tanto el histograma como los descriptores que se obtienen a partir del mismo pueden ser utilizados en diversas operaciones que permitan por ejemplo la clasificación de una imagen por su intensidad en oscura, media o clara utilizando para esto la media.

3.2.3.1 DIVISIÓN POR NIVELES

La división por niveles consiste en la agrupación de los datos de la muestra a partir de intervalos definidos, clasificados para un posterior procesamiento específico aplicado a cada grupo.

Los intervalos de valores tomados para la determinación de valores oscuros, medios y claros se muestran en la Tabla 3.1.

Intensidad	Valores
Oscuro	0-100
Medio	100-160
Claros	160-255

Tabla 3.1 Tabla de valores para la delimitación entre valores oscuros, medios y claros
Fuente: Elaboración propia

Como se mencionó en el capítulo anterior la media y desviación estándar determinan medidas globales sobre la iluminación y el contraste de la imagen, evaluando estos valores es posible determinar si se deben realizar operaciones adicionales sobre la imagen especialmente para la mejora de las mismas mediante operaciones de realce, permitiendo una imagen de mejor calidad y más fácil de procesar en las siguientes etapas.

Por ejemplo, en la figura 3.3 se observa una imagen de intensidad clara ya que la mayoría de su punto más alto se encuentra en el valor 250 por lo cual podríamos situarla en el nivel de intensidad clara.

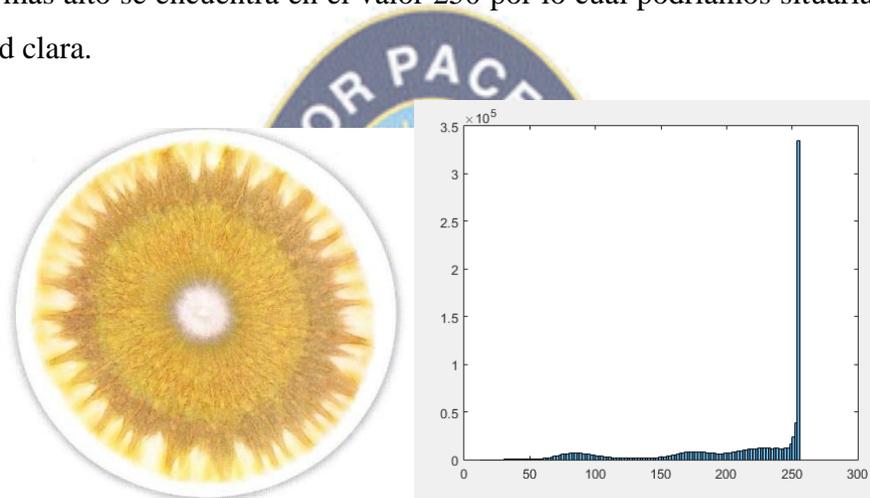


figura 3.3 Cromatograma con su respectivo histograma

Fuente: Elaboración propia

3.2.3.2 AJUSTE DEL HISTOGRAMA

De acuerdo al valor del promedio de intensidad se utilizara una combinación diferente de modificadores del contraste y modificadores de histograma, la primera operación permitirá una mejora de la imagen que dependerá únicamente de la intensidad en general de la imagen, en el caso de imágenes con niveles oscuros se aplicaran las funciones logarítmica, raíz cubica y raíz cuadrada, para mejorar la iluminación, en el caso de imágenes sobreexpuestas o muy claras se utilizaran las funciones cuadrática y cubica con el fin de disminuir el brillo, el objetivo de aplicar modificadores es obtener imágenes con una uniformidad relativa en cuanto a intensidad y a partir de esto se obtendrá una imagen en escala de grises para proceder a la segmentación.

3.3 SEGMENTACIÓN

Esta etapa se ha dividido en dos tareas para mejorar el resultando en el tiempo de ejecución; en la primera se extrae el área de interés mediante umbralización y en la segunda se segmenta la imagen mediante el algoritmo k-means este proceso se puede apreciar en la figura 3.4



Figura 3.4 Secuencia de procesos de la segmentación

Fuente: Elaboración Propia

3.3.1 EXTRACCIÓN DEL ÁREA DE INTERÉS

Dado que el procedimiento de segmentación propiamente dicho es computacionalmente intenso, la primera parte del procedimiento propuesto consiste en aislar las zonas externa, intermedia, interna y central y excluir del análisis la zona periférica que no es trascendente en el proceso de segmentación.

Este procedimiento se realizará a través de un proceso de umbralización sencillo. En el cual se determina el área de interés de la imagen según la intensidad, siendo el área de interés los pixeles de la imagen con una intensidad de media a oscura con valores menores a 190 y mayores a 100.

En la figura 3.5 se observa que al aplicar el algoritmo de umbralización básico se logra eliminar la zona periférica del cromatograma, obteniendo así la imagen deseada.

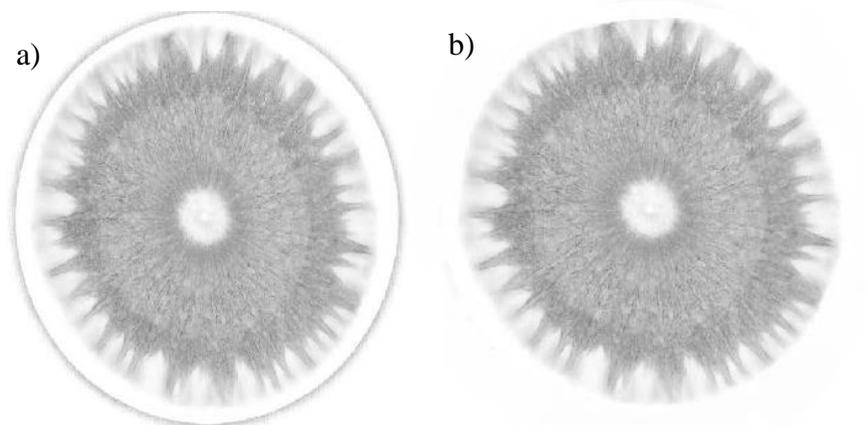


Figura 3.5 Imagen en escala de grises (izquierda); imagen umbralizada (derecha)

Fuente: Elaboración propia

3.3.2 SEGMENTACIÓN DEL ÁREA DE INTERÉS

El algoritmo de umbralización filtra de la imagen original del fondo, dejando solo la región concerniente al cromatograma. Es sobre esta región de la imagen que aplicaremos el algoritmo de segmentación de técnica de clustering, K-means, con el fin de obtener las zonas externa, intermedia, interna y central.

Así, la imagen del cromatograma debidamente pre-procesado servirá de entrada al algoritmo de segmentación, que asignará a cada pixel de esta imagen un clúster correspondiente de cuatro clusters posibles (zona externa, zona intermedia, zona interna y zona central).

Posteriormente se procede a separar la segmentación en las distintas partes en la que se descompone, aplicando máscara R (rojo), máscara G (verde) y máscara B (azul) a cada uno de los componentes, esto resultará en k imágenes a color, cada uno representando a una zona del área de interés

Donde la primera imagen representa la zona periférica la que como mencionamos en una sección anterior será desechada ya que no brinda ningún tipo de información para la

representación y el reconocimiento de la imagen, la segunda imagen obtenida define la zona intermedia de cromatograma, la tercera imagen determina la zona central u ombligo del cromatograma; la cuarta imagen delimita a la zona interna del cromatograma, finalmente la última imagen obtenida representa a la zona intermedia; cabe señalar que para todos los componentes del cromatograma se obtendrá la información del color.

3.4 PLANTEAMIENTO DEL ALGORITMO

Inicialmente se tiene como entrada al algoritmo la representación de la imagen, que como ya mencionamos anteriormente estará en formato BMP del cual se obtendrá su representación como estructura de datos; en este caso una matriz bidimensional de tamaño 640*480 cuyo almacenamiento está definido en *Bottom-left*, en el modelo RGB; esto se puede observar más gráficamente en la figura 3.6.

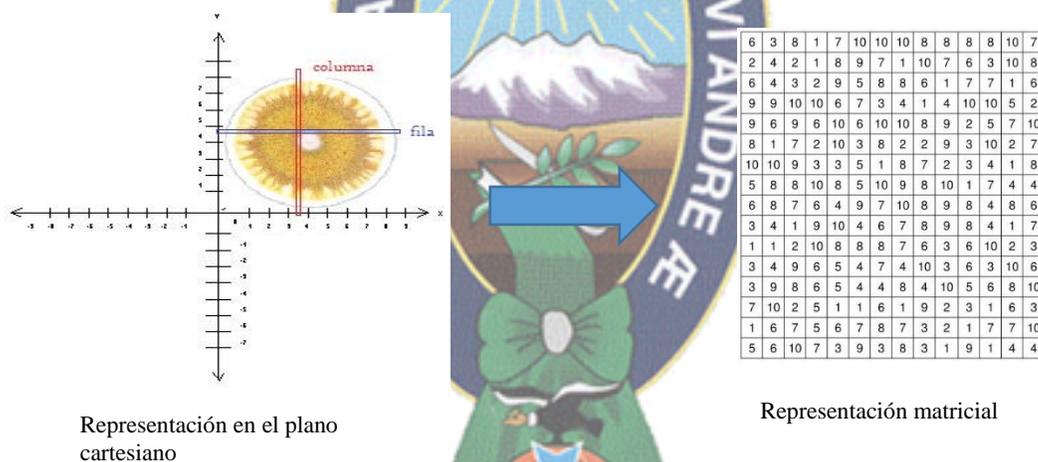


Figura 3.6 Entrada y salida de la representación de imagen
Fuente: elaboración propia

Ya teniendo la representación matricial de la imagen procedemos a calcular su histograma y respectivos descriptores los cuales nos servirán para realizar la operación de división por niveles y ajuste de histograma. El flujo de este proceso de muestra en la Figura 3.7.

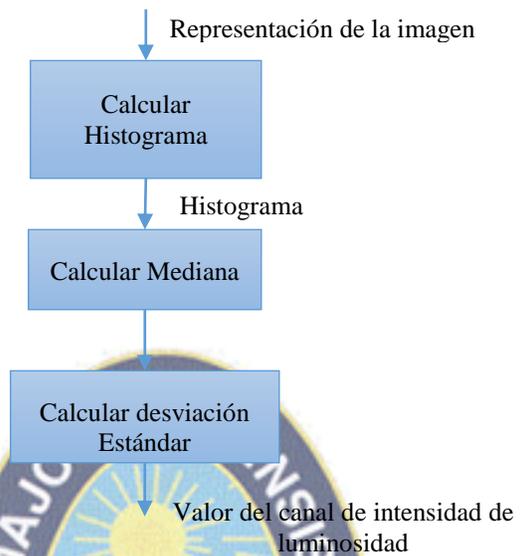


Figura 3.7 Diagrama de Flujo cálculo de descriptores
Fuente: Elaboración propia

Teniendo el valor de la media del canal de intensidad luminosa se procede a determinar el nivel de luminosidad ya sea claro, medio u oscuro según los valores definidos en la tabla 3.1, el flujo de este proceso se observa en la Figura 3. 8

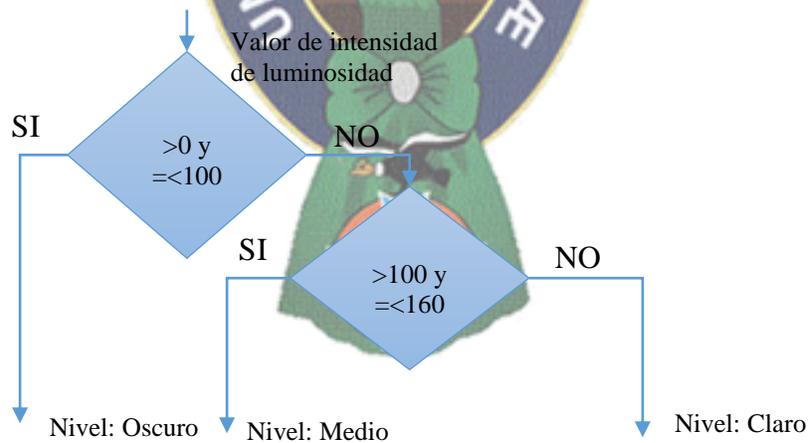


figura 3.8 Diagrama de flujo del proceso de División por niveles
Fuente: Elaboración propia

Habiendo obtenido el nivel de luminosidad procedemos a realizar el ajuste de histograma, los valores de entrada se limitan a tres valores: clara, media u oscura, obtenidos del anterior proceso, de modo que con esta información se proceda a realizar una de las siguientes operaciones:

a) Intensidad Oscura (Intensidad <100)

Caso 1: **Intensidad [0,85[**

- Se Aplica la ecualización logarítmica
- Obtener imagen es encala de grises
- Realizar una expresión tangente con valor alfa= 0,70

Caso 2: **Intensidad [85,100]**

- Se aplica una ecualización de raíz cubica
- Obtener una imagen es escala de grises
- Realizar una expansión tangente con un valor alfa=0,75

b) Intensidad Media (**$100 \leq \text{Intensidad} \leq 160$**)

- Aplicar una ecualización raíz cuadrada
- Obtener una imagen en escala de grises
- Realizar una expansión tangente con un valor alfa=0,80

c) Intensidad Clara (**$160 < \text{Intensidad}$**)

Caso 1: Intensidad [**160,180[**

- Aplicar una ecualización cuadrática
- Obtener una imagen en escala de grises

Caso 2: Intensidad [**180,255]**

- Obtener una imagen en escala de grises
- Realzar una expansión tangente con un valor alfa=0,85

La ecualización logarítmica se realiza con el fin de realzar los pixeles oscuros sin saturar pixeles brillantes, de la misma forma la ecualización de raíz cubica solo que en un distinto

rango. La ecualización de raíz cuadrada se realiza con el fin de mantener realzar tanto pixeles brillantes como oscuros. Finalmente, la ecualización cuadrática que pertenece a la ecualización de tipo exponencial se realiza con el fin de realzar los pixeles brillantes sin saturar pixeles oscuros.

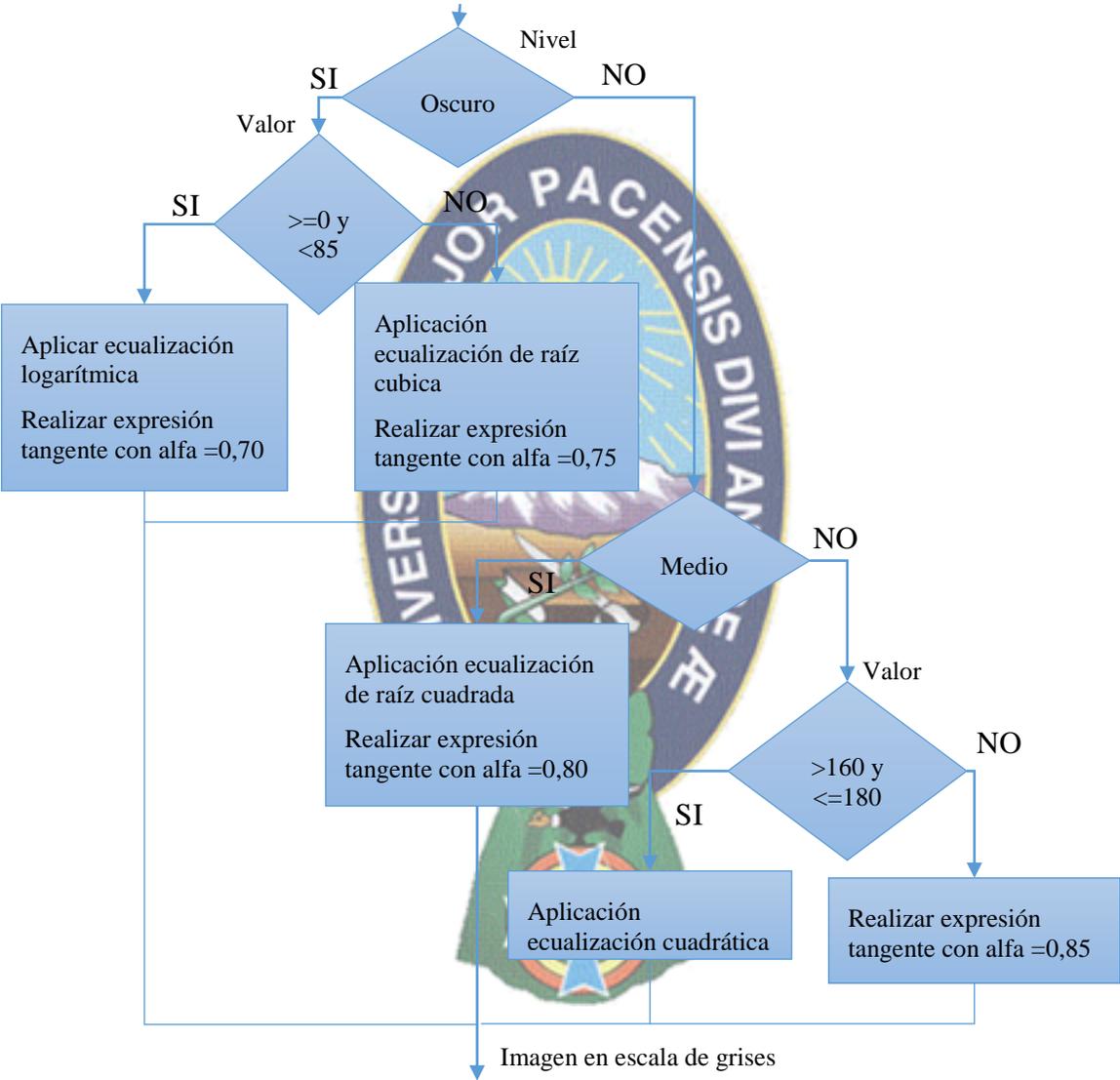


figura 3.9 Diagrama de flujo Ajuste de Histograma
Fuente: Elaboración propia

Teniendo la imagen en escala de grises el siguiente paso es la segmentación de la imagen, lo cual se llevará a cabo, como ya se mencionó anteriormente en dos etapas, la primera en la que se extrae el área de interés aplicando umbralización, que para el caso de imágenes de cromatografía se observó que se adquieren mejores resultados seleccionando los pixeles de interés con valor de intensidad menores a 195. El flujo de este proceso se observa en la Figura 3.10

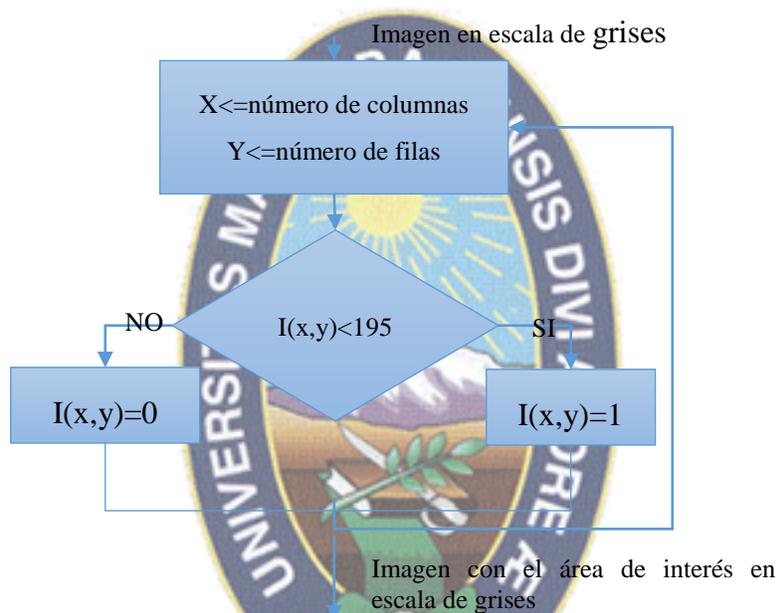


figura 3.10 Diagrama de flujo del proceso de umbralización
Fuente: Elaboración Propia

La umbralización simple se lleva a cabo con el fin de excluir pixeles cuyo nivel de intensidad es mayor igual a 195, es decir los pixeles que cumplan con este criterio recibirán el nivel máximo de intensidad, es decir, 255; por el contrario, los que no cumplan con dicho criterio, recibirán el nivel mínimo de intensidad, es decir 0. Posterior a este procedimiento se obtiene una imagen cuya matriz de representación estará constituida únicamente por dos valores '0' y '255', a la cual se la transformará en una imagen en escala de grises, que es más sencilla de segmentar.

Finalmente, se procede a la segmentación de las partes constituyentes de la imagen de cromatografía mediante el algoritmo K-means, para el cual se recibe como entrada una imagen en escala de grises que contiene el área de interés.

En esta etapa, para el caso particular de imágenes de cromatografía de suelos, habiendo notado una mejor segmentación, se han establecido, los siguientes parámetros:

- **K = 4**
- **Medida de similitud: Color**
- **Inicialización de centroide: Aleatorio (Método Forgy)**
- **Calculo de distancia: Distancia Euclidea**

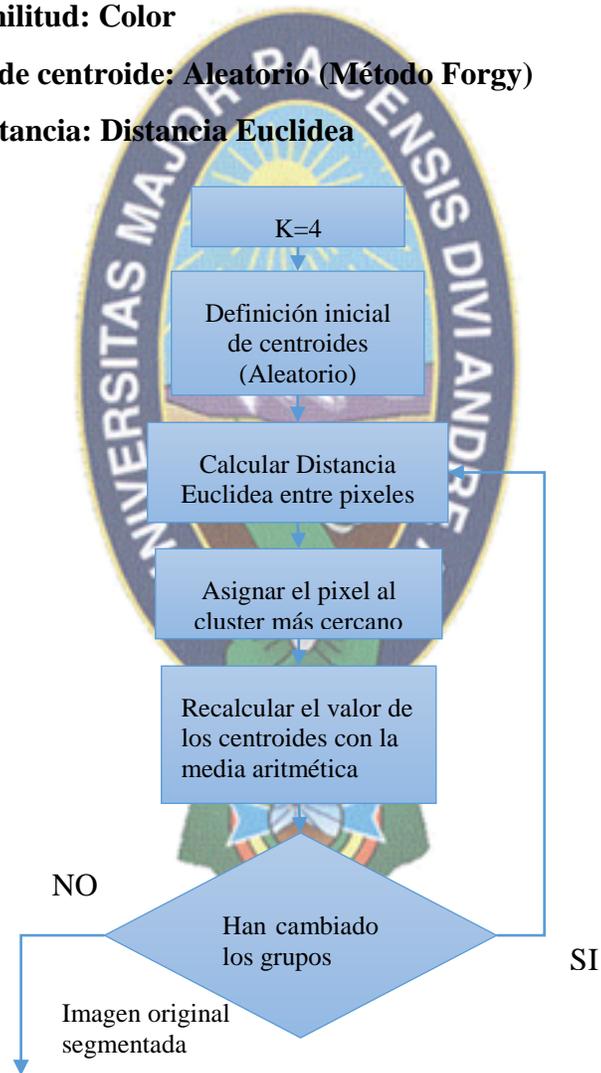


figura 3.11 Diagrama de flujo de la segmentación mediante K-means
Fuente: Elaboración propia

La figura 3.11 muestra el flujo de la segmentación por k-means

3.5 EVALUACIÓN DE LA SEGMENTACIÓN

La métrica para a evaluación de la segmentación que ofrece el algoritmo propuesto es únicamente el tiempo que es medido desde el inicio del proceso hasta que los segmentos son adquiridos. Cabe notar que no se evaluara la calidad de la segmentación ya que no se cuenta con una base de datos de segmentación referencia para imágenes de cromatografía.

Para este fin se tomó como ejemplo a dos imágenes que fueron adquiridas de la elaboración de cromatogramas del análisis de suelos en la localidad de Sapecho Provincia Nor Yungas del Departamento de La Paz, las cuales fueron capturadas con la cámara de un teléfono móvil de gama media con una resolución de 8 megapíxeles, almacenadas en formato BMP, las cuales se observan en la figura 3.12.

La imagen del cromatograma 3.12 a) tomada de un cultivo de naranja convencional con dos años de recuperación con biofertilizantes y caldos minerales, representa para esta investigación una referencia de un suelo sano, por otra parte, la imagen del cromatograma 3.12 b) tomada de un suelo con cultivo de naranja con un impacto negativo de la mala aplicación de urea, representa para esta investigación un referente de un suelo un estado intermedio o de recuperación.

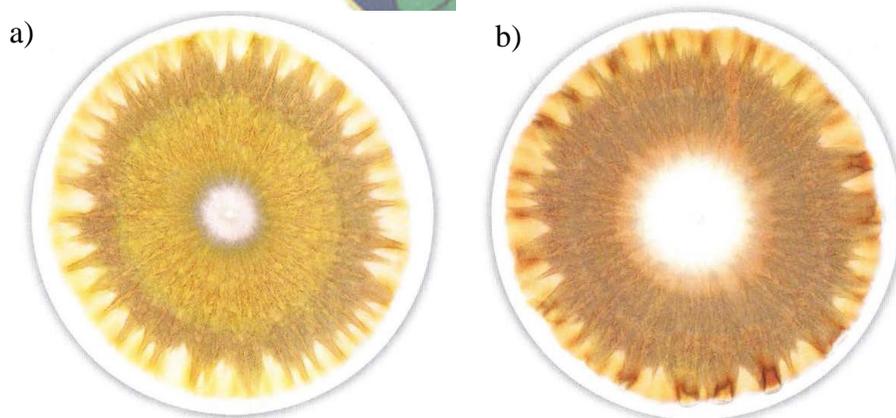
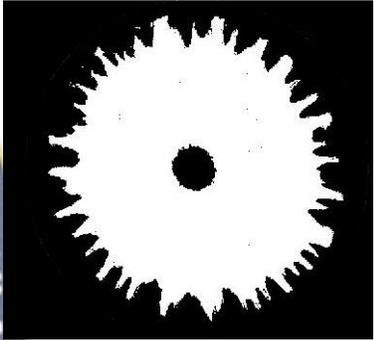
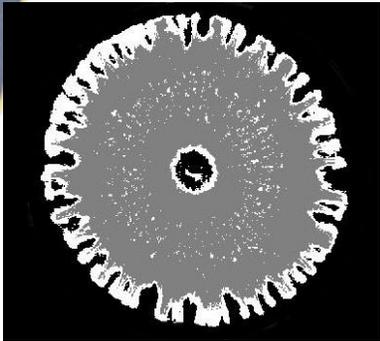
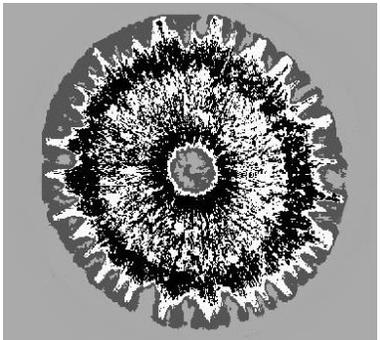


figura 3.12 Cromatogramas de dos diferentes tipos de suelo
Fuente: Elaboración propia

A continuación, en la tabla 3.2 se muestra las imágenes resultantes de la segmentación de la imagen 3.12 a) más el tiempo de ejecución resultantes de las iteraciones aplicando el algoritmo a la imagen resultante de la umbralización hecha anteriormente, con distintos valores de k empezando desde k=2 hasta k=6.

Estos resultados se muestran en la tabla 3.3.

Valor de k	Tiempo de ejecución (seg)	Segmentación obtenida
K=2	2,780	
K=3	2,986	
K=4	3,025	

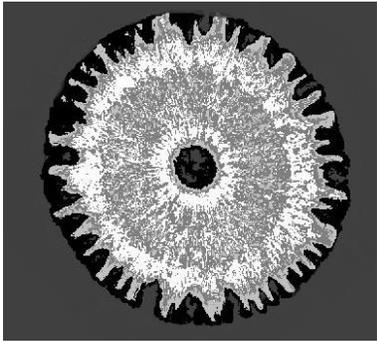
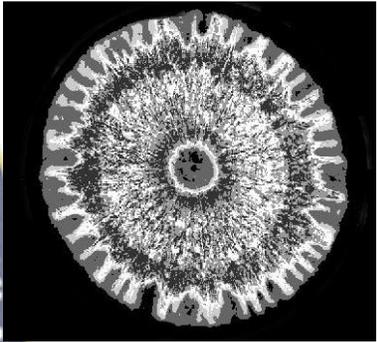
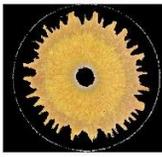
K=5	3,250	
K=6	3,350	

Tabla 3.2 Resultados de segmentación con algoritmo k-means para distintos valores de k

Fuente: Elaboración propia

Posteriormente se procedió a dividir cada segmento para las cinco imágenes obtenidas para cada valor de k, es en este paso en donde se logra distinguir más claramente para cuál de los valores de K se logra una mejor segmentación de las zonas de interés del cromatograma. En la tabla 3.3 se muestra cada segmento asignado a una zona si es que este existiera en el valor de k asignado.

Valor de k	Zona central	Zona interna	Zona intermedia	Zona externa
2			No se distingue	Unida a la zona centra

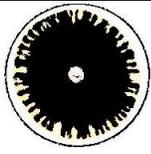
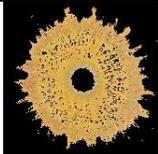
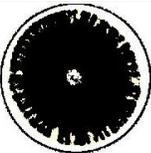
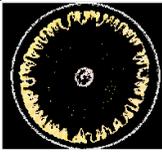
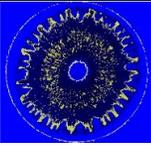
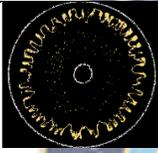
3				Unida a la zona intermedia
4				
5		Genera una imagen que no sirve para la representación		
6			Genera imágenes que no son de utilidad para la representación	

Tabla 3.3 Comparación de segmentos adquiridos según el valor de k

Fuente: Elaboración Propia

Como se puede apreciar en la tabla 3.3 la iteración cuando el valor de $k=4$ obtiene una mejor segmentación, que logra separar las cuatro zonas de interés del cromatograma.

3.6 REPRESENTACIÓN

Como ya se mencionó en el anterior capítulo en esta etapa del PDI se debe tomar la decisión si los datos de salida de la segmentación deben ser representados como límites, como una completa región o una complementación de ambas representaciones. Para el caso particular de la segmentación de imágenes de cromatogramas, tomando en cuenta los aspectos que deben a ser evaluados para la interpretación de los mismos.

Para el caso de esta investigación se determina representar los segmentos como una completa región, ya que para el reconocimiento solo se tomará en cuenta el color de los píxeles presentes en cada segmento hallado, el cual nos brindará información básica sobre el estado de los suelos.

3.7 RECONOCIMIENTO

En esta etapa del procesamiento se utilizarán las imágenes obtenidas en la etapa de segmentación dividiéndolas en sus diferentes componentes como podemos observar del ejemplo en la figura 3.13

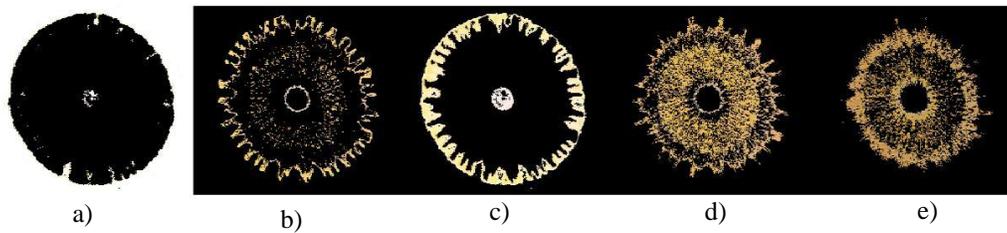


Figura 3.13 Componentes de la segmentación de un cromatograma

Fuente: Elaboración propia

Para esta etapa se tiene como base de información las principales coloraciones definidas en Restrepo y Pinheiro (2012) en donde indica seis tonalidades que indican una buena salud del suelo, estas son:

- Amarillo
- Dorado
- Rojizo
- Anaranjado
- Café claro
- Tonalidades verdosas

Por otra parte, también indica que la combinación entre cafés claros o muy oscuros reflejan un estado intermedio.

También define seis tonalidades que reflejan la mala salud del suelo, estas son:

- Negro
- Ceniza
- Café muy oscuro

- Lilas o violeta
- Gris
- Tonalidades azuladas

3.8 DISEÑO DEL PROTOTIPO

3.8.1 DECLARACIÓN DE PROPÓSITOS

La aplicación tiene como propósito servir de prototipo de prueba para brindar un diagnóstico del estado de un suelo, posterior al procesamiento digital de imágenes a través de la segmentación de la imagen del cromatograma del suelo en cuestión.

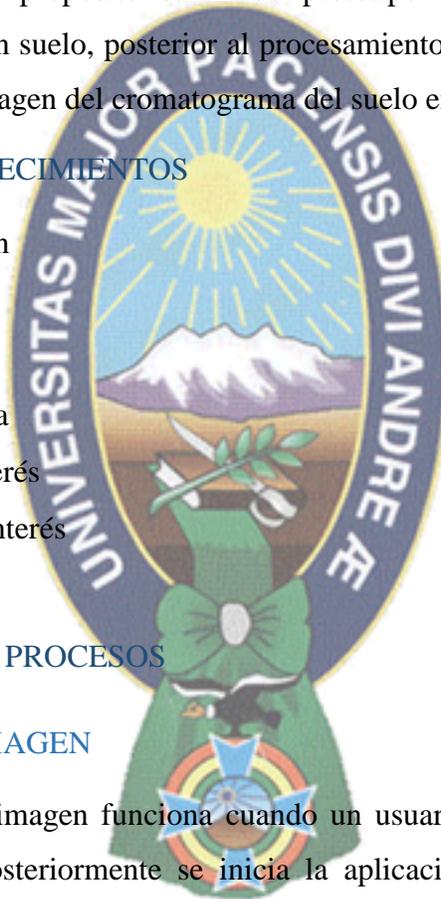
3.8.2 LISTA DE ACONTECIMIENTOS

- Captura de la imagen
- Calcular histograma
- División por niveles
- Ajuste de histograma
- Hallar el área de interés
- Segmentar área de interés
- Reconocimiento

3.8.3 DESCRIPCIÓN DE PROCESOS

3.8.3.1 CAPTURA DE IMAGEN

El proceso de Captura de imagen funciona cuando un usuario realiza una solicitud para “tomar una fotografía”, posteriormente se inicia la aplicación de cámara que se tenga instalada en el teléfono móvil, el usuario podrá tomar la fotografía, posterior a ello podrá determinar según su criterio si desea tomar otra fotografía, de lo contrario acepta la fotografía capturada e inmediatamente se procesa el cálculo del histograma.



3.8.3.2 CALCULAR HISTOGRAMA

Para iniciar este proceso la entrada es la imagen capturada por el usuario, el proceso inicia al leer los datos en el buffer de imagen procediendo a calcular la distribución de frecuencias para cada canal y almacenando los mismos en histograma por canal, con base a en los cálculos de histogramas se procederá al cálculo de descriptores de imagen, como ser la media, la desviación estándar, los cuales dan una idea general de la forma y la posición de los datos al interior del histograma. El dato más importante es la media la cual representa el valor de intensidad en un determinado canal (rojo, verde, azul o luminosidad).

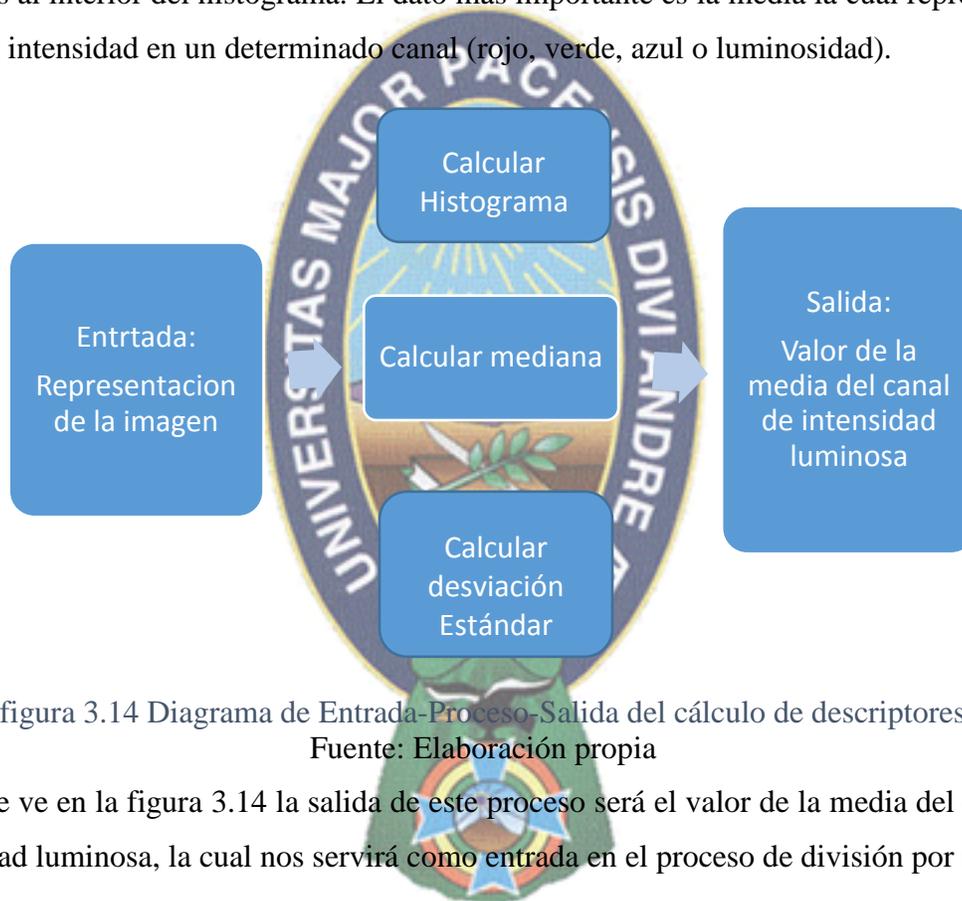


figura 3.14 Diagrama de Entrada-Proceso-Salida del cálculo de descriptores
Fuente: Elaboración propia

Como se ve en la figura 3.14 la salida de este proceso será el valor de la media del canal de intensidad luminosa, la cual nos servirá como entrada en el proceso de división por niveles.

3.8.3.3 DIVISIÓN POR NIVELES

Este proceso tiene como entrada el valor de la media del canal de intensidad luminosa, que es tomado para determinar la clasificación de la imagen como clara, media u oscura. Esto es importante ya que a partir de esa clasificación se proporciona un tratamiento diferenciado para cada nivel de intensidad.



Figura 3.15 Diagrama Entrada-Proceso- Salida de la división por niveles

Fuente: Elaboración propia

Como se observa en la figura 3.15 la salida de este proceso es una variable que contenga la clasificación de la imagen lo cual será pasado como parámetro al proceso de Ajuste de histograma.

3.8.3.4 AJUSTE DE HISTOGRAMA

Este proceso tiene como entrada la clasificación de la imagen, los valores de entrada se limitan a tres valores: clara, media u oscura, de modo que con esta información se proceda a realizar operaciones de ajuste de histograma para obtener una imagen de mejor calidad más fácil de procesar



figura 3.16 Diagrama Entrada-Proceso- Salida de la división por niveles

Fuente: Elaboración propia

En la figura 3.16 se puede observar que la salida de este proceso es una imagen en escala de grises de la imagen original, la cual servirá de entrada para el proceso de umbralización de la imagen.

3.8.3.5 HALLAR ÁREA DE INTERÉS

La entrada de este proceso será una imagen escala de grises obtenida del proceso anterior, este proceso tendrá como objetivo segmentar la imagen en dos clústeres uno que será llamado el área de interés y el restante que será depurado realizando una umbralización buscando valores de intensidad de media a oscura con valores menores a 190 y mayores a 100.



Figura 3.17 Diagrama Entrada-Proceso- Salida de la umbralización

Fuente: Elaboración propia

Ya obtenida la segmentación por umbralización se tendrá una imagen de la cual multiplicaremos su representación matricial con la representación matricial de la imagen original en escala de grises, en otros términos, se aplicará la segmentación por umbralización como mascarará para la imagen original en escala de grises, de esta manera se habrá separado el área de interés. La salida de este proceso será una imagen en escala de grises, como se ve en la figura 3.17, la cual servirá como entrada al siguiente proceso.

3.8.3.6 SEGMENTAR EL ÁREA DE INTERÉS

La entrada a este proceso será una imagen en escala de grises que solo contendrá el área de interés, dicha imagen será procesada con el algoritmo k-means con un valor inicial para $k=4$ de tal manera que se logre obtener cuatro clusters que representaran a cada una de las zonas del cromatograma, de este proceso se obtendrá una imagen segmentada.

Lo siguiente a este proceso será separar los segmentos en imágenes diferentes para ello se aplicará a cada clúster una tres mascarar:

- a) La imagen de entrada original en su componente R (rojo).
- b) La imagen de entrada original en su componente G (verde).

c) La imagen de entrada original en su componente B (azul).

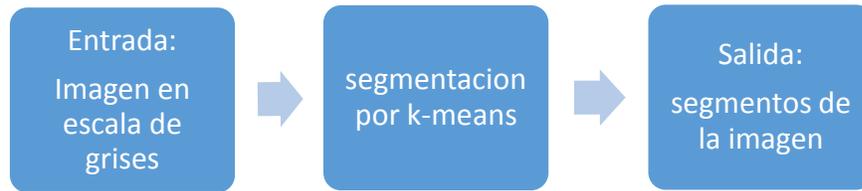


figura 3.18 Diagrama Entrada-Proceso- Salida de la segmentación
Fuente: Elaboración propia

De tal manera que se obtenga cada uno de los cuatro segmentos en cuatro imágenes en formato RGB, las cuales se constituirán en las salidas de este proceso como se puede observar en la figura 3.18.

3.8.3.7 RECONOCIMIENTO

Las entradas para este proceso son las cuatro imágenes segmentadas las cuales serán interpretadas como ya se había indicado en el punto 3.6, para finalmente obtener el diagnóstico del estado del suelo que se está analizando.

3.8.4 DIAGRAMAS DE FLUJO DE DATOS

A continuación, se presentan los diagramas de flujo de datos para el prototipo de la aplicación para diagnóstico del estado del suelo mediante reconocimiento de patrones.

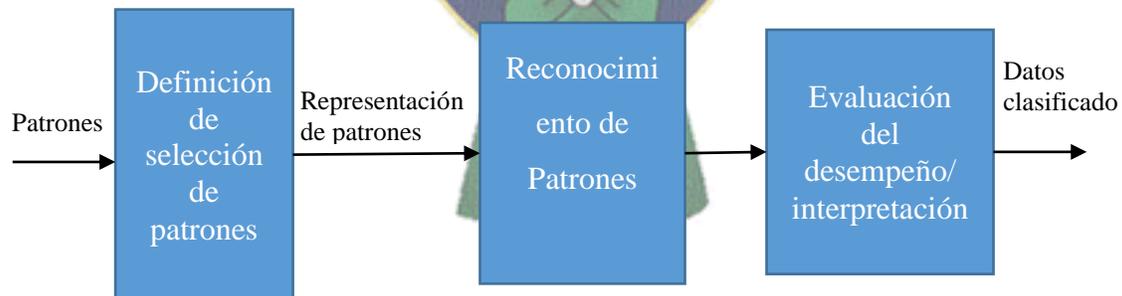


figura 3.19 Diagrama en bloques del proceso de reconocimiento de patrones
Fuente: elaboración propia

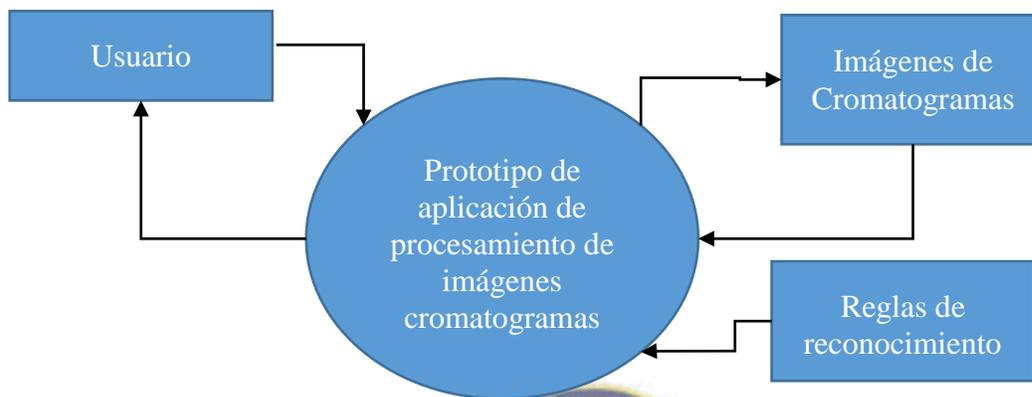


figura 3.20 Prototipo de aplicación de procesamiento de imágenes de cromatogramas.
Fuente: elaboración propia

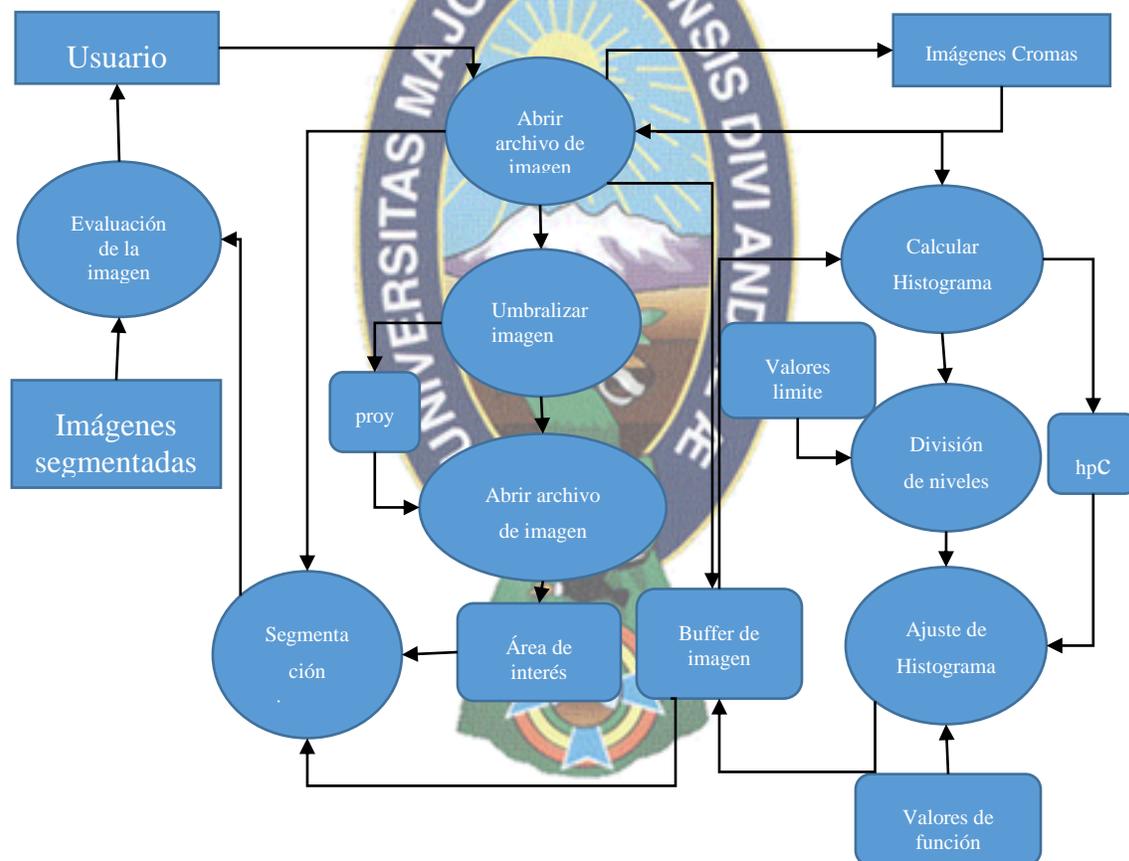


figura 3.21 Prototipo de aplicación de procesamiento de imágenes de cromatogramas.
Diagrama de primer nivel
Fuente: Elaboración propia

3.8.5 IMPLEMENTACIÓN DE ALGORITMOS

3.8.5.1 ALGORITMO DE HISTOGRAMA Y DESCRIPTORES

A continuación, se muestra el algoritmo implementado para el cálculo del histograma.

```
import java.awt.Color;
import java.awt.image.BufferedImage;
public class Histograma {
    /**
     * Calculamos la media de una variable Color
     * @param color del cual se quiere obtener la media
     * @return entero con el valor de la media
     */
    private int calcularMedia(Color color){
        int mediaColor;
        mediaColor=(int)((color.getRed()+color.getGreen()+color.getBlue())/3);
        return mediaColor;
    }
    /**
     * Devuelve el histograma de la imagen.
     * @param imagen BufferedImage de la cual se quiere obtener el histograma
     * @return Devuelve una variable int[5][256], donde el primer campo[0] corresponde
     * al canal Rojo, [1]=verde [2]=azul [3]=alfa [4]=escala grises
     */
    public int[][] histograma(BufferedImage imagen){
        Color colorAuxiliar;
        /*Creamos la variable que contendrá el histograma
        El primer campo [0], almacenará el histograma Rojo
        [1]=verde [2]=azul [3]=alfa [4]=escala grises*/
        int histogramaReturn[][]=new int[5][256];
        //Recorremos la imagen
        for( int i = 0; i < imagen.getWidth(); i++ ){
            for( int j = 0; j < imagen.getHeight(); j++ ){
                //Obtenemos color del píxel actual
                colorAuxiliar=new Color(imagen.getRGB(i, j));
                //Sumamos una unidad en la fila roja [0],
                //en la columna del color rojo obtenido
                histogramaReturn[0][colorAuxiliar.getRed()]+=1;
                histogramaReturn[1][colorAuxiliar.getGreen()]+=1;
                histogramaReturn[2][colorAuxiliar.getBlue()]+=1;
                histogramaReturn[3][colorAuxiliar.getAlpha()]+=1;
                histogramaReturn[4][calcularMedia(colorAuxiliar)]+=1;
            }
        }
        return histogramaReturn;
    }
}
```

3.8.5.2 IMPLEMENTACIÓN DEL ALGORITMO K-MEANS

Para implementar el algoritmo se elaboraron 2 clases que representan a los objetos Cluster y Punto. Estas clases se muestran a continuación:

```
public class Cluster {
    public List points;
    public Point centroid;
    public int id;
    //Crea un Nuevo cluster
    public Cluster(int id) {
        this.id = id;
        this.points = new ArrayList();
        this.centroid = null;
    }
    public List getPoints() {
        return points;
    }
    public void addPoint(Point point) {
        points.add(point);
    }
    public void setPoints(List points) {
        this.points = points;
    }
    public Point getCentroid() {
        return centroid;
    }
    public void setCentroid(Point centroid) {
        this.centroid = centroid;
    }
    public int getId() {
        return id;
    }
    public void clear() {
        points.clear();
    }
    public void plotCluster() {
        System.out.println("[Cluster: " + id+"]");
        System.out.println("[Centroid: " + centroid + "]");
        System.out.println("[Points: \n");
        for(Point p : points) {
            System.out.println(p);}
        System.out.println("]");
    }
}
```

```

public class Point {
    private double x = 0;
    private double y = 0;
    private int cluster_number = 0;

    public Point(double x, double y)
    {
        this.setX(x);
        this.setY(y);
    }
    public void setX(double x) {
        this.x = x;
    }
    public double getX() {
        return this.x;
    }
    public void setY(double y) {
        this.y = y;
    }
    public double getY() {
        return this.y;
    }
    public void setCluster(int n) {
        this.cluster_number = n;
    }
    public int getCluster() {
        return this.cluster_number;
    }
    //Calculates the distance between two points.
    protected static double distance(Point p, Point centroid) {
        return Math.sqrt(Math.pow((centroid.getY() - p.getY()), 2) +
Math.pow((centroid.getX() - p.getX()), 2));
    }
    //Crear un punto random
    protected static Point createRandomPoint(int min, int max) {
        Random r = new Random();
        double x = min + (max - min) * r.nextDouble();
        double y = min + (max - min) * r.nextDouble();
        return new Point(x,y);
    }
    protected static List createRandomPoints(int min, int max, int
number) {
        List points = new ArrayList(number);
        for(int i = 0; i < number; i++) {
            points.add(createRandomPoint(min,max));
        }
        return points;
    }
    public String toString() {
        return "("+x+", "+y+")";
    }
}

```

3.8.6 INTERFAZ DEL PROTOTIPO

El prototipo de la aplicación de diagnóstico del estado de los suelos para dispositivos móviles, fue desarrollado en el IDE Android Studio 2.2.2 desarrollado con una interfaz gráfica diseñada en función del análisis y diseño ya presentados, opera bajo sistema operativo Android a partir del API 15.

La interfaz del prototipo se puede observar de manera detallada en el anexo 1.



PRUEBA DE HIPÓTESIS

4.1 INTRODUCCIÓN

En este capítulo inicialmente se presenta una breve descripción de la prueba de rachas de Wald Wolfowitz para dar una idea global del funcionamiento de la misma, posteriormente se desglosa la solución estadística aplicada al prototipo de la aplicación diseñado para el diagnóstico del estado de los suelos, para la comprobación de la hipótesis planteada y finalmente se interpreta los resultados obtenidos.

La evaluación del prototipo se realizó en la comunidad de Sapecho con pobladores del lugar, tratando de abarcar todos los posibles casos existentes.

4.2 PRUEBA DE RACHAS DE WALD-WOLFOWITZ

La prueba de rachas sirve para determinar si una muestra de observaciones es o no aleatoria, es decir, para determinar si las observaciones de una determinada secuencia son independientes entre sí. En una serie temporal, por ejemplo, las observaciones no son aleatorias: lo que ocurre con una observación cualquiera depende, generalmente, de las características de la observación anterior. En una muestra aleatoria, por el contrario, debemos esperar que lo que ocurre con una observación cualquiera sea independiente de las características de la anterior (y de la siguiente).

El concepto de racha hace referencia a una secuencia de observaciones de un mismo tipo. Supongamos que lanzamos una moneda al aire 10 veces y que obtenemos el siguiente resultado:

CCCXCCXXXC

Tendremos 5 rachas: CCC, X, CC, XXX y C. A simple vista, el resultado obtenido parece aleatorio. Pero si en lugar de ese resultado hubiéramos obtenido este otro:

CCCCCXXXXX (2 rachas)

Resultaría fácil ponernos de acuerdo en que la secuencia obtenida no parece aleatoria. Como tampoco parece aleatoria una secuencia con demasiadas rachas:

CXCXCXCXCX (10 rachas)

Pues bien, la prueba de las rachas permite determinar si el número de rachas (R) observado en una determinada muestra de tamaño n es lo suficientemente grande o lo suficientemente pequeño como para poder rechazar la hipótesis de independencia (o aleatoriedad) entre las observaciones⁴.

Para obtener el número de rachas es necesario que las observaciones estén clasificadas en dos grupos exhaustivos y mutuamente exclusivos (variable dicotómica). Si no lo están, deberemos utilizar algún criterio (mediana, media, moda, etc.) para hacer que lo estén (variable dicotomizada).

Una vez clasificadas las n observaciones en dos grupos (de tamaños n_1 y n_2), se utiliza una tipificación⁵ del número de rachas (R) para contrastar la hipótesis de aleatoriedad o independencia:

$$Z = \frac{R - E(R)}{\sigma_R}$$

El estadístico Z se distribuye según el modelo de probabilidad normal $N(0, 1)$.

⁴ Conviene no confundir la hipótesis de aleatoriedad con la hipótesis de bondad de ajuste estudiada a propósito de la prueba binomial.

Obtener 5 caras y 5 cruces al lanzar una moneda 10 veces es un resultado que se ajusta perfectamente a la hipótesis de equiprobabilidad ($\pi \text{ cara} = \pi \text{ cruz} = 0,5$), pero si las 5 caras salen al principio y las cinco cruces al final, esto haría dudar de la hipótesis de independencia o aleatoriedad.

⁵ Si el tamaño muestral es menor que 50, el estadístico Z se obtiene utilizando la *corrección por continuidad* de la siguiente manera:

- Si $R - E(R) < -0,5$, se suma 0,5 a R . Es decir: $Z = [R + 0,5 - E(R)]/\sigma_R$.
- Si $R - E(R) > 0,5$, se resta 0,5 a R . Es decir: $Z = [R - 0,5 - E(R)]/\sigma_R$.
- Si $|R - E(R)| \leq 0,5$, $Z = 0$.

Para muestras aleatorias, la distribución de probabilidad de R tiende hacia la normal, a medida que N_1 y N_2 , se van agrandando, de tal manera que:

$$R \rightarrow N(E[R]) * \sqrt{Var[R]}$$

Siendo:

$$E[R] = \frac{2 * N_1 * N_2 + N}{N}$$

$$S(R) = \frac{2 * N_1 * N_2 (2 * N_1 * N_2 - N_1 - N_2)}{(N_1 + N_2)^2 * (N_1 + N_2 - 1)}$$

4.3 ESTADO DE LA HIPÓTESIS

Para realizar la prueba de la hipótesis se ha tomado como espacio de muestra a 18 pobladores de la Comunidad de Sapecho–Nor yungas previamente capacitados en la elaboración de cromatogramas, a los cuales se les pidió interpretar un cromatograma, inicialmente sin el software que se tiene como prototipo y posteriormente con el software.

Evaluando su capacidad de diagnóstico en escala del 0 al 10 donde:

- 0 no realiza un diagnóstico y,
- 10 se realizó el diagnóstico al 100%

Nº	Con Software (C)	Sin Software (S)
1	10	0
2	2	8
3	7	3
4	8	2
5	10	0
6	5	5

7	3	7
8	10	0
9	8	2
10	5	5
11	7	3
12	10	0
13	3	7
14	8	2
15	4	6
16	10	0
17	8	2
18	3	7

Tabla 4.1 Muestras tomadas para evaluación del prototipo
Fuente: Elaboración Propia

Paso1: Se colocan los puntajes $N_1 + N_2$ en una sola serie ordenada

0	0	0	0	0	2	2	2	2	2	2	3	3	3	3
N	N	N	N	N	Y	N	N	N	N	N	N	Y	N	Y
3	4	5	5	5	5	6	7	7	7	7	7	7	8	8
Y	Y	Y	N	Y	N	N	Y	N	Y	N	N	N	N	Y
8	8	8	10	10	10	10	10							
Y	Y	Y	Y	Y	Y	Y	Y							

Tabla 4.2 Prueba de Rachas
Fuente: Elaboración Propia

Paso2: SE DETERMINA EL NÚMERO DE RACHAS

En este caso son 14 rachas

Paso3: HIPÓTESIS

Se tiene la hipótesis nula:

H_0 : El uso del algoritmo de agrupamiento k-means para la segmentación de imágenes de cromatografía de suelos permite realizar un diagnóstico del estado de los mismos con una confiabilidad de por lo menos 80%.

Se tiene como Hipótesis Alternativa la negación de la Hipótesis nula:

H_1 : El uso del algoritmo de agrupamiento k-means para la segmentación de imágenes de cromatografía de suelos permite realizar un diagnóstico del estado de los mismos con una confiabilidad menor al 80%.

Paso 4: NIVEL DE SIGNIFICACIÓN

Para una muestra de 18 personas el nivel de significancia llegaría a ser 0.05

$$N = 36$$

$$\alpha = 0.05$$

$$N_1 = 18$$

$$N_2 = 18$$

$$R = 14$$

Paso 5: REGLA DE DECISIÓN

La tabla muestra que para los valores de $N_1 = 18$ y $N_2 = 18$, una R de 14 es significativa al nivel de 0.05.

$$E[R] = \frac{2 * N_1 * N_2 + N}{N}$$



$$E[R] = \frac{2 * 18 * 18 + 36}{36}$$

$$= 19$$

$$S(R) = \frac{2 * N_1 * N_2(2 * N_1 * N_2 - N_1 - N_2)}{(N_1 + N_2)^2 * (N_1 + N_2 - 1)}$$

$$S(R) = \frac{2 * 18 * 18(2 * 18 * 18 - 18 - 18)}{(18 + 18)^2 * (18 + 18 - 1)}$$

$$= \frac{648(612)}{(36)^2 * (35)}$$

$$= \frac{396576}{1296 * 35}$$

$$= \frac{396576}{45360}$$

$$= 8.743$$

Como la regla de decisión al 80% de confianza, no se rechazará la hipótesis nula de aleatoriedad H_0 si el número de rachas se encuentra en el intervalo:

$$[E(R) - S(R)] - [E(R) + S(R)]$$

En el presente caso tomamos los valores $[10.257 - 27.743]$, prueba que $R = 14$, pertenece al intervalo, de esta manera se acepta la hipótesis, por tan podemos afirmar:

H_0 : El uso del algoritmo de agrupamiento k-means para la segmentación de imágenes de cromatografía de suelos permite realizar un diagnóstico del estado de los mismos con una confiabilidad de por lo menos 80%.

CONCLUSIONES Y RECOMENDACIONES

5.1 INTRODUCCIÓN

En este capítulo se describen los principales hallazgos realizados como producto de la presente investigación, y mencionamos algunas recomendaciones para posibles investigaciones futuras.

5.2 CONCLUSIONES

Se ha logrado plantear un algoritmo para la segmentación de imágenes de cromatografía de suelos, el cual está basado en el algoritmo de agrupamiento k-means y tiene como medida de similitud el color; el cual permite apoyar en el diagnóstico del estado de los suelos.

El algoritmo nos permite segmentar las imágenes de cromatografía de suelos de tal manera que se logran distinguir las zonas central, interna, intermedia y externa, que son las zonas de interés para la interpretación de un cromatograma de suelos, para posteriormente llegar a un diagnóstico del estado de los mismos.

Se pretende que al ser implementado este algoritmo se puedan elaborar herramientas que incentiven a los pobladores a realizar el análisis del estado de sus suelos previo al cultivo y así determinen el mejor tipo de uso de sus suelos agrícolas y con esto evitar el deterioro de los mismos y de la misma forma disminuir el costo de dicho análisis y el tiempo de su elaboración.

De acuerdo con los resultados de la prueba de hipótesis podemos concluir que el algoritmo planteado para la segmentación de imágenes de cromatografía de suelos permite apoyar en el diagnóstico del estado de los mismos con una confiabilidad de por lo menos 80%.

La elaboración del prototipo de la aplicación móvil para apoyar en el diagnóstico del estado de los suelos, constituye una herramienta útil para lograr este objetivo, el cual, al ser usada por pobladores de distintas comunidades, representa la disminución del uso de recursos

monetarios en traslado de muestras a laboratorios, también en cuanto al tiempo en el que se procesan las mismas.

La elaboración del prototipo de la aplicación móvil ha sido de gran ayuda para apoyar en el diagnóstico del estado de los suelos, constituye una herramienta útil para lograr este objetivo, el cual puede ser usada por pobladores de distintas comunidades.

5.3 RECOMENDACIONES

Deberían explorarse otros dominios de imágenes que ofrezcan mejores condiciones para el desempeño del algoritmo propuesto, así como evaluar diversos niveles de ruido. Asimismo, debería hacerse un mayor análisis para verificar el impacto de los parámetros en la calidad de la segmentación.

También, el número de clústeres/segmentos en el algoritmo actual son ingresados de forma manual. Deberían explorarse mecanismos que permitan la obtención automática y óptima de este parámetro. Asimismo, pueden realizar otras tareas de pre-procesamiento sobre la imagen de entrada que faciliten el proceso de segmentación.

Se debería realizar la segmentación mediante otros algoritmos u otras variaciones del algoritmo k-means para evaluar cuál de ellos logra una mejor calidad en la segmentación.

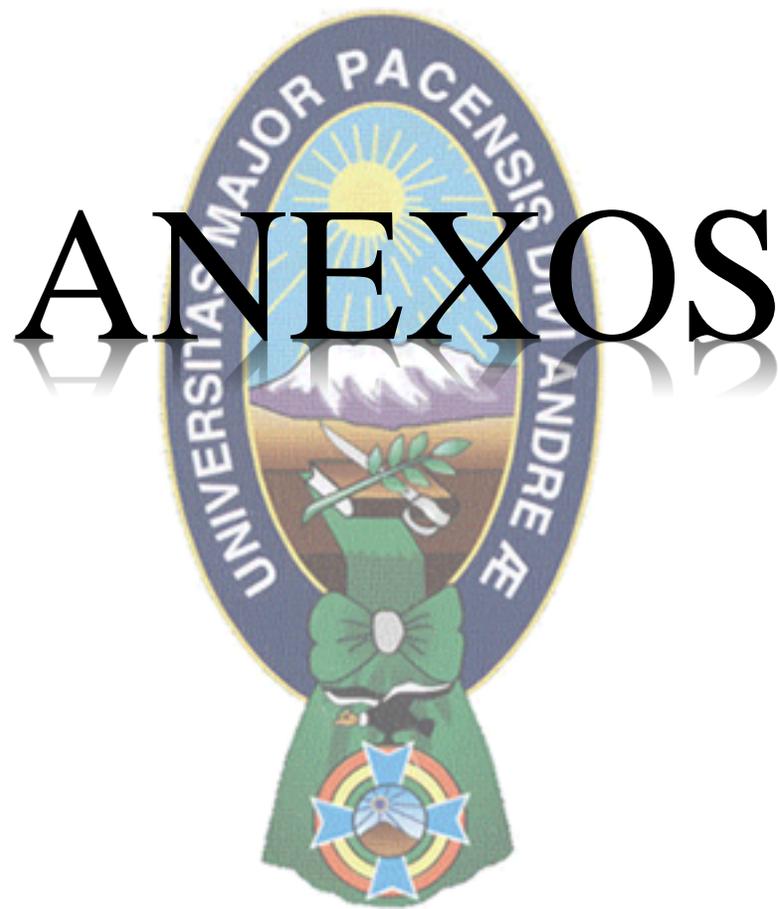
Así mismo para la evaluación de la segmentación de imágenes de cromatografía de suelos, con otros algoritmos se recomienda utilizar las imágenes obtenidas en los resultados de esta investigación como referencia de la calidad de la segmentación ya que actualmente no se cuenta con una base de datos con resultados estándares de este tipo de imágenes.

Por otra parte, también se debería experimentar la segmentación con otra medida de similitud como ser Manhattan o Mahalanobis para reconocer patrones en cuanto a textura, evolución radial entre otros y así brindar un diagnóstico completo del estado de los suelos.

BIBLIOGRAFÍA

- [1] Alvarado J. (2012), *Procesamiento y Análisis de Imágenes Digitales*, Costa Rica: Instituto Tecnológico de Costa Rica
- [2] Angulo J. y Iñigo R. (1986), *Visión Artificial por Computador*, Paraninfo S.A.: Magallanes
- [3] Baldivieso J. y Zapata F. (2006), *Sistema de evaluación y administración integrada del uso de la tierra*, La Paz: Viceministerio de Desarrollo Alternativo.
- [4] Comas D.S. y Meschino G.J. (2015). *Segmentación de Imágenes Mediante Reconocimiento de Patrones*, Buenos Aires: Escuela y Workshop Argentino En ciencias de las Imágenes.
- [5] Flores O. (2014). *Sistema de diagnóstico de la retinopatía diabética mediante imágenes digitales*, La Paz: Autor.
- [6] Gavidia J.C. (2014). *Segmentación de imágenes médicas mediante algoritmos de colonias de hormigas*, San Miguel: Universidad Católica de Perú.
- [7] Gonzalez R.C. y Woods R.E. (2002). *Digital imagen processing*, New Jersey: Upper Saddle River.
- [8] Guevara G. y Sánchez O. (2015, noviembre). *Segmentación de imágenes a color mediante técnicas de agrupamiento de datos empleando los algoritmos k-means y c-means*, *Boletín UPIITA*, 51. Obtenida el 13 de junio de 2016, <https://www.boletin.upiita.ipn.mx/index.php/ciencia/639-cyt-numero-51/1192>
- [9] Herrera V. (2014). *Implementación del algoritmo de Fourier para identificar especies vegetales adoptándolo a dispositivos móviles Android*, La Paz: Autor
- [10] Jain A.K. Murty M.N. y Flynn P.J., (1999). "Data Clustering: A Review", *ACM Computing Surveys*
- [11] Martin M. (2013), *Procesamiento Digital de Imágenes*,
- [12] Mejía J. (2005). *Apuntes de Procesamiento digital de Imágenes*, San Luis Potosí: Facultad de Ingeniería UALP

- [13] Meschino G.J. Andrade R.A.E. y Ballarin V.L. (2006a). *Reconocimiento de tejidos en imágenes cerebrales de Resonancia Magnética a través de valores de verdad de predicados difusos*, Mecánica Computacional
- [14] Meschino G.J. Andrade R.A.E. y Ballarin V.L. (2006b). *Reconocimiento de tejidos en imágenes cerebrales de Resonancia Magnética a través de valores de verdad de predicados difusos*, Mecánica Computacional.
- [15] Meschino G.J. Andrade R.E. y Ballarin V.L. (2008). *A framework for tissue discrimination in Magnetic Resonance brain images based on predicates analysis and Compensatory Fuzzy Logic*, International Journal of Intelligent Computing in Medical Sciences and Image Processing IC-MED.
- [16] Moler E.G. Ballarin V.L. y Gonzalez M. (1999). *Características Estadísticas, Espectrales y Morfológicas para Clasificación de Texturas: Un Análisis Comparativo*, VIII RPIC Reunión de Trabajo en Procesamiento de la Información y Control.
- [17] Pinto L. (2015). *Análisis de la aplicación de algoritmos de K-means y Continuous Max-Flow a la segmentación de imágenes en color*, Sevilla: Autor
- [18] Restrepo J. y Pinheiro S. (2011), *Cromatografía: Imágenes de vida y destrucción del suelo*. Colombia: Feiba S. A.
- [19] Tarqui B. (2013). *Cuantificación y detección de alteraciones morfológicas de glóbulos rojos en imágenes digitales microscópicas*, La Paz: Autor.
- [20] Toscano G. Pereira O. y Arzola J. (2010), *Segmentación de imágenes digitales mediante técnicas de clustering complementadas con técnicas de crecimiento de regiones*, Cuba: Universidad de las Ciencias Informáticas
- [21] Torres J. R. (2013). *Interpretación de cromatogramas de suelos*.
- [22] Ugarte C. (2015). *Manejo sostenible de suelos*, Obtenida el 12 de junio de 2016, en <http://lapatriaenlinea.com/?nota=2>



ANEXOS

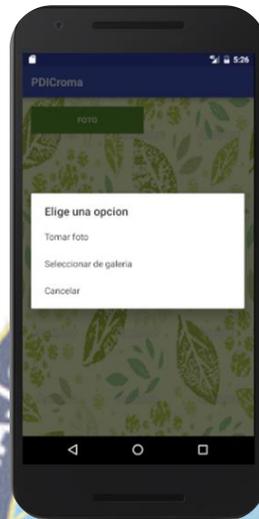
INTERFAZ DEL PROTOTIPO

Al ingresar a la aplicación por primera vez se pedirán permisos de para utilizar los recursos de cámara y para acceder a la galería de fotos y archivos. Esto se aprecia en las siguientes figuras:



Posteriormente al permitir estos permisos la primera vez y en el inicio para las siguientes veces se observara la pantalla de inicio que consta de un unicamente de un boton

denominado “FOTO”, el cual al precionar nos abra un dialogo con opciones de tomar foto o abrir la galeria para seleccionar una o cancelar. Esto se obserba en la siguiente imagen:



Al elegir la opción “Tomar foto” se abrirá aplicación que permite tomar fotos del dispositivo, después de capturar la imagen nos encontramos con una pantalla para aceptar la fotografía o cancelarla. De manera similar la opción “Seleccionar de galería”, nos abre la aplicación de galería de imágenes del dispositivo, seleccionamos la foto la cual, de igual manera, podemos aceptar o cancelar. Obsérvese en la siguiente imagen.

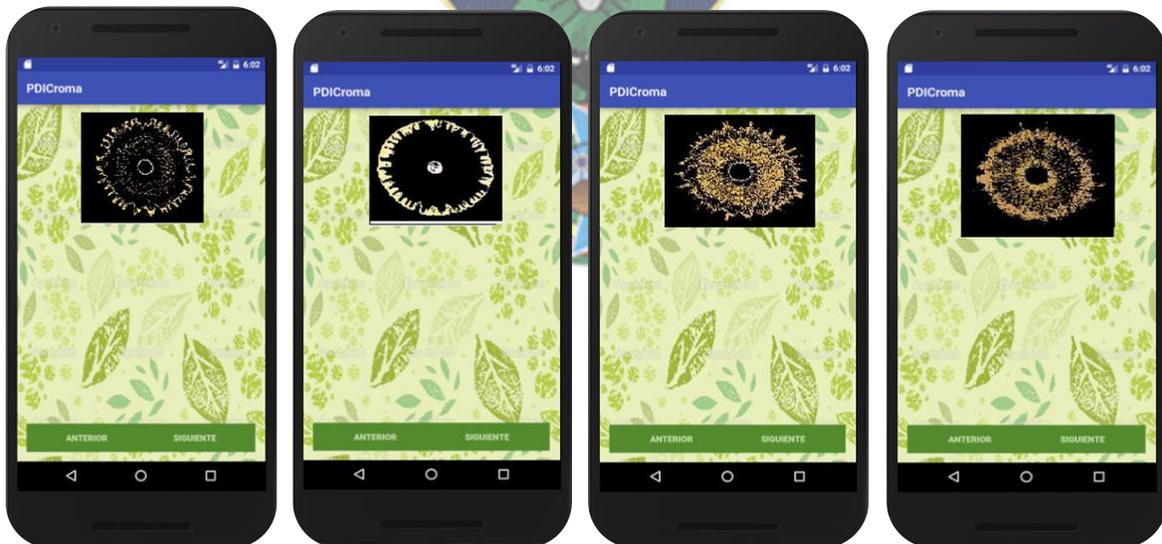


Al presionar “Aceptar” se procede a la segmentación de la imagen, al finalizar el proceso se muestra los resultados de este en una nueva pantalla, caso contrario regresa a la pantalla inicial.



En la pantalla anterior se muestra un diagnostico cualitativo del estado del suelo basándose únicamente en la representación de los colores que presenta el cromatograma.

El botón “Ver Zonas” nos llevara a una pantalla que se podrán ver los segmentos obtenidos por el algoritmo de segmentación.



DOCUMENTACIÓN

