



Universidad Veracruzana **UNIVERSIDAD VERACRUZANA**

TESIS DE MAESTRÍA

**Clasificación de poblaciones nativas de frijol
(*Phaseolus vulgaris* L.) utilizando aprendizaje
automático sobre histogramas de color**

Presenta:
José Luis Morales Reyes

Director de tesis:
Dr. Héctor Gabriel Acosta Mesa

Codirector:
Dra. Elia Nora Aquino Bolaños

*Una tesis presentada en cumplimiento de los requisitos
para obtener el grado de Maestro en Inteligencia Artificial*

en el

CIIA
Centro de Investigación en Inteligencia Artificial

24 de enero de 2018

Universidad Veracruzana

Resumen

Centro de Investigación en Inteligencia Artificial

Clasificación de poblaciones nativas de frijol (*Phaseolus vulgaris L.*) utilizando aprendizaje automático sobre histogramas de color

Maestro en Inteligencia Artificial

José Luis Morales Reyes

En el presente trabajo se plantea una propuesta para la solución del problema de clasificación de poblaciones nativas de frijol basada en color para la identificación por variedad y color del grano. Actualmente esta medición se realiza a través de un espectrofotómetro para sólidos, el cual solamente permite la medición puntual de color sobre el grano, para solucionar la limitante se propone un método que considera la totalidad de la distribución de color sobre la superficie de los granos de las distintas poblaciones nativas de frijol. La metodología aquí propuesta hace uso de una cámara fotográfica y un prototipo de iluminación que mantiene luz constante mediante una caja de luz. Este ambiente de iluminación controlado reduce los reflejos especulares y mantiene una iluminación uniforme. La segmentación de imagen se realiza mediante un algoritmo de crecimiento de región por color. A partir de esto se calculan histogramas de dos y tres dimensiones de los espacios de color HSI y CIE $L^*a^*b^*$, representados como probabilidades conjuntas. Estas características son utilizadas por el algoritmo de aprendizaje automático de k vecinos más cercanos con distancia ponderada. Los resultados de este trabajo muestran que éste método es robusto comparado con los resultados generados con el espectrofotómetro, para identificación por variedad local con histogramas de dos y tres dimensiones obteniendo 75.19 % hasta 80.46 % y para identificación por color del grano 91.61 % hasta 94.44 %, con el espectrofotómetro lográndose 51.54 % de precisión para identificar variedades locales y 69.99 % para identificar por color del grano. Este método aporta la base para nuevas investigaciones basadas en color para estimar propiedades químicas relacionadas con el color y resulta más rápido y económico en comparación con el espectrofotómetro para sólidos.

Agradecimientos

Las muestras de frijol analizadas en el presente estudio forman parte del proyecto CONACYT-2015-01-119: **“Alimentos y productos de maíz y frijol pigmentados con alto valor nutricional para las comunidades oaxaqueñas de alta marginación”**, dirigido por el Dr. José Luis Chávez Servia, investigador del Instituto Politécnico Nacional, CIIDIR, Oaxaca.

Mi agradecimiento al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca económica 712056 para la realización de este proyecto.

Al Centro de Investigación en Inteligencia Artificial de la Universidad Veracruzana por darme la oportunidad de ser estudiante y haber logrado un meta más en la vida.

Al Instituto de Ciencias Básicas de la Universidad Veracruzana por permitirme hacer estancia de investigación.

Le agradezco la confianza a mis asesores Dr. Héctor Gabriel Acosta Mesa y Dra. Elia Nora Aquino Bolaños por haberme dado la oportunidad de desarrollar nuestra tesis profesional y darme la oportunidad de aprender cosas nuevas.

A mis sinodales Dra. Elia Nora Aquino Bolaños, Dr. Homero Vladimir Ríos Figueroa y Dr. Nicandro Cruz Ramírez que me brindaron su tiempo y amabilidad para leer este trabajo y darme sus observaciones.

Dedicatorias

Le dedico y le agradezco a Dios que me dio la oportunidad de vivir y darme una hermosa familia.

A mis Padres Eduardo Morales Moreno y Alicia Reyes Jiménez por su incansable amor, apoyo incondicional y que gracias a ellos la oportunidad de tener una excelente educación en el transcurso de mi vida.

A mi esposa Delfina Contreras Garrido por su amor y comprensión de todos aquellos fines de semana que dedique a mi formación y que comparto contigo este objetivo logrado.

A mis hijos para que sepan que los logros se ganan a base de esfuerzos y los sueños se pueden hacer realidad, que la educación da el sentido a la vida y que este documento sea testimonio de mi gran amor hacia ellos.

Participación en el VII Foro Nacional de Divulgación Científica y Tecnológica 2017

La Universidad Veracruzana
Otorga el presente

Reconocimiento

a

JOSÉ LUIS MORALES REYES

Por su destacada participación como:

Ponente

En el marco del

VII Foro Nacional de Divulgación Científica y Tecnológica,

FODICYT 2017

Celebrado los días 25, 26 y 27 de octubre de 2017, en las instalaciones de la Facultad de Ingeniería Mecánica y Eléctrica, en la ciudad de Xalapa, Veracruz.

FIME XALAPA

ESTECS

FIME XALAPA

TenarisTamsa

POLIFLEX

ADO

nuvoil

Laboratorio de Ribología

Dra. Rosario Aldana Franco
Presidente de Logística

Dr. Elym Jesús Álvarez Sánchez
Presidente del Comité Técnico

FIGURA 1: Reconocimiento FODICYT 2017.

Índice general

Resumen	III
Agradecimientos	V
Dedicatorias	VII
1. Introducción	1
1.1. Motivación	4
1.2. Justificación	7
1.3. Hipótesis	7
1.4. Objetivos	7
1.4.1. Objetivo General	7
1.4.2. Objetivos Específicos	7
2. Antecedentes	9
3. Metodología	19
3.1. Materiales y métodos	20
3.1.1. Poblaciones de granos de frijol	20
3.1.2. Software	20
3.2. Modelos de color	20
3.2.1. Espacio de color	20
3.3. El modelo de color RGB	20
3.4. El espacio de color HSI	21
3.4.1. Conversión de espacios de color RGB a HSI	22
3.5. El espacio de color CIE L*a*b*	23
3.5.1. Conversión del espacio de color RGB a CIE L*a*b*	23
3.6. Adquisición de imágenes	24
3.6.1. Prototipo de iluminación	24
3.6.2. Caja de luz para fotografía	25
3.6.3. Cámara fotográfica	26
3.6.4. Proceso de adquisición de imágenes	27
3.7. Proceso de segmentación	28
3.7.1. Crecimiento de región	28
3.7.2. Obtención de los pesos para el algoritmo de segmentación	30
3.7.3. Selección de granos para entrenamiento y prueba	30
3.8. Histogramas	32
3.8.1. Histograma 1D	33
3.8.2. Distribución de probabilidad conjunta	34

3.8.3.	Histograma de 2 dimensiones	36
3.8.4.	Histogramas en 3 dimensiones	37
3.8.5.	Extracción de histogramas	38
3.9.	Paleta de colores	42
3.10.	Aprendizaje automático	44
3.10.1.	K vecinos más cercanos	44
3.10.2.	Algoritmo K vecinos más cercanos con distancia ponderada	45
3.10.3.	Clasificación de histogramas con K vecinos más cercanos con distancia ponderada	46
4.	Resultados	47
4.1.	Resultados de clasificación espectrofotómetro	47
4.2.	Resultados de clasificación por variedad local utilizando histogramas	49
4.2.1.	Resultados con el espacio de color HSI	49
4.2.2.	Resultados con el espacio de color CIE L*a*b*	49
4.2.3.	Resultados con el espacio de color HSI y CIE L*a*b*	50
4.3.	Resultados de clasificación por color del grano utilizando histogramas	51
4.3.1.	Resultados con el espacio de color HSI	51
4.3.2.	Resultados con el espacio de color CIE L*a*b*	51
4.3.3.	Resultados con los espacios de color HSI y CIE L*a*b*	52
5.	Discusión	61
6.	Conclusiones y trabajo futuro	65
A.	Georreferencia de las 54 poblaciones nativas de frijol	67
B.	54 Poblaciones nativas de frijol	71
C.	Descripción de color para las 54 poblaciones de frijol	79
D.	Resultado de prueba estadística Kruskal–Wallis	81
D.1.	Resultados gráficos estadísticos de clasificación de poblaciones nativas de frijol por su variedad local	81
D.1.1.	Utilizando histogramas de 2 dimensiones del espacio de color HSI	81
D.1.2.	Utilizando histogramas de 3 dimensiones del espacio de color HSI	81
D.1.3.	Utilizando histogramas de 2 dimensiones del espacio de color CIE L*a*b*	82
D.1.4.	Utilizando histogramas de 3 dimensiones del espacio de color CIE L*a*b*	82
D.1.5.	Utilizando la unión de los histogramas de 2 dimensiones de los espacios de color HSI con CIE L*a*b*	82
D.1.6.	Utilizando la unión de los histogramas de 3 dimensiones de los espacios de color HSI con CIE L*a*b*	83
D.2.	Resultados gráficos estadísticos de clasificación de poblaciones nativas de frijol por color del grano	83
D.2.1.	Utilizando histogramas de 2 dimensiones del espacio de color HSI	83
D.2.2.	Utilizando histogramas de 3 dimensiones del espacio de color HSI	84
D.2.3.	Utilizando histogramas de 2 dimensiones del espacio de color CIE L*a*b*	84
D.2.4.	Utilizando histogramas de 3 dimensiones del espacio de color CIE L*a*b*	84

D.2.5. Utilizando la unión de los histogramas de 2 dimensiones de los espacios de color HSI con CIE L*a*b* 85

D.2.6. Utilizando la unión de los histogramas de 3 dimensiones de los espacios de color HSI con CIE L*a*b* 85

Referencias **87**

Índice de figuras

1.	Reconocimiento FODICYT 2017	VIII
1.1.	Medición con espectrofotómetro	2
1.2.	Variación de color en las poblaciones de frijol	2
1.3.	Variación en la tonalidad en frutos y semillas	3
1.4.	Medición con espectrofotómetro	4
1.5.	Histograma granos de frijol color negro	4
1.6.	Histograma granos de frijol de dos tonalidades	5
1.7.	Población 02 y Población 33	5
1.8.	Histograma granos de diversos colores	6
2.1.	5 tipos de frijol	10
2.2.	Gráfica comparativa de clasificación por variedad local	11
2.3.	Gráfica comparativa de clasificación por variedad local	12
2.4.	3 variedades de frijol	13
2.5.	16 tipos de frijol	14
2.6.	67 tipos de frijol	15
3.1.	Metodología propuesta para la caracterización	19
3.2.	Espacio de color RGB	21
3.3.	Espacio de color HSI	22
3.4.	Espacio de color CIELAB	23
3.5.	Esquema prototipo de iluminación	25
3.6.	Prototipo de iluminación	26
3.7.	Granos de frijol de la población 22	26
3.8.	Pequeño conjunto de poblaciones de frijol	27
3.9.	Ejemplo de segmentación de la población 59	29
3.10.	Ejemplo de selección de granos para prueba y entrenamiento	32
3.11.	Representación de una imagen	32
3.12.	Notación matriz compacta y tradicional	33
3.13.	Histograma	34
3.14.	Histograma probabilidad conjunta	36
3.15.	Ejemplo de histograma en 2-D	37
3.16.	Paleta de colores	43
4.1.	Gráfica comparativa de clasificación por variedad local	50
4.2.	Gráfica comparativa de clasificación por color del grano	52
D.1.	Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 2 dimensiones del espacio de color HSI	81

D.2. Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 3 dimensiones del espacio de color HSI	81
D.3. Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 2 dimensiones del espacio de color CIE L*a*b*	82
D.4. Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 3 dimensiones del espacio de color CIE L*a*b*	82
D.5. Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 2 dimensiones de los espacios de color HSI con CIE L*a*b*	82
D.6. Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 3 dimensiones de los espacios de color HSI con CIE L*a*b*	83
D.7. Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 2 dimensiones del espacio de color HSI	83
D.8. Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 3 dimensiones del espacio de color HSI	84
D.9. Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 2 dimensiones del espacio de color CIE L*a*b*	84
D.10. Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 3 dimensiones del espacio de color CIE L*a*b*	84
D.11. Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 2 dimensiones de los espacios de color HSI con CIE L*a*b*	85
D.12. Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 3 dimensiones de los espacios de color HSI con CIE L*a*b*	85

Índice de cuadros

2.1. Comparativa de trabajos relacionados con la identificación de variedades de frijol, se despliega las características utilizadas para la clasificación.	17
3.1. Configuración de disparo de la cámara fotográfica.	27
3.2. Taxonomía de color para poblaciones nativas de frijol.	43
4.1. Resultados de clasificación variedad local con valores de medición de espectrofotómetro	48
4.2. Table caption text	48
4.3. Resultados de clasificación por variedad local utilizando histogramas de 2 y 3 dimensiones del espacio de color HSI	54
4.4. Resultados de clasificación por variedad local histogramas de 2 y 3 dimensiones CIE L*a*b*	55
4.5. Resultados de clasificación histogramas 2 y 3 dimensiones HSI con CIE L*a*b*	56
4.6. Resultados de clasificación de poblaciones nativas de frijol por color del grano histogramas 2 y 3 dimensiones HSI	57
4.7. Resultados de clasificación por color del grano con histogramas 2 y 3 dimensiones del espacio de color CIE L*a*b*	58
4.8. Resultados de clasificación por color del grano con histogramas 2 y 3 dimensiones de los espacios de color HSI con CIE L*a*b*	59
A.1. Georreferencia de las 54 poblaciones nativas de frijol	67
C.1. Taxonomía de color.	79

Lista de algoritmos

1.	Algoritmo de crecimiento de región por color	28
2.	Generación de máscaras de segmentación para granos de entrenamiento y prueba	31
3.	Generación de histogramas en 2-D de HSI a 256 bins	38
4.	Generación de histograma 2-D de a^* y b^* de CIE $L^*a^*b^*$	39
5.	Generación de histogramas en 3-D de HSI a 128 bins	40
6.	Generación de histograma 3-D de CIE $L^*a^*b^*$ a 128 bins	41
7.	El algoritmo básico K vecinos más cercanos para aproximar a las funciones de valor discreto.	45

Capítulo 1

Introducción

El frijol (*Phaseolus vulgaris* L.) desde hace 8000 años, evolucionó como grano silvestre y fue domesticado en la regiones altas de latinoamérica [12], es un miembro de la familia *Fabacea*, del genero *Phaseolus* y de la especie *Phaseolus vulgaris* L. [15]. La importancia del frijol junto con el maíz son la base de la alimentación, ya que es buena fuente de proteínas y de carbohidratos, rico en fibra y buena fuente de vitaminas [16], en México hay un gran número de variedades que en muchos casos se cultivan para el consumo particular del agricultor [29], y existen cientos de variedades locales domesticadas y poblaciones silvestres de frijol [3].

México es el centro de origen, diversificación y domesticación del frijol basado en evidencia arqueológica, etnobotánica, morfológica, bioquímica y evidencia genética e isoenzimática[9][6], las semillas de frijol contienen una amplia variedad de componentes nutricionales benéficos para la salud como proteínas, fibra, polifenoles, flavonoides, entre otros componentes relacionados con la actividad antioxidante, que previene efectos cardiovasculares o enfermedades crónico degenerativas, como cáncer, obesidad y diabetes, así como otras condiciones relacionadas con el síndrome metabólico, triglicéridos y colesterol [11]. Existe un especial interés en la clasificación de poblaciones nativas de frijol, para disponer de un método que permita reconocer a las distintas poblaciones de frijol tanto por su variedad local como por su coloración.

El color del frijol generalmente se mide utilizando un espectrofotómetro para sólidos, el cual es un instrumento de análisis de alta precisión del color principalmente en laboratorios y en aplicaciones de investigación y desarrollo [20], este dispositivo se conecta a una computadora para ser controlado y con lo que se permite activar la orden de medición, tiene una pequeña abertura circular de 8 mm donde se coloca el objeto a medir, de una medición de la superficie se reporta tres valores correspondientes a cada canal del espacio de color CIE $L^*a^*b^*$ que es un descriptor de color ampliamente utilizado en la medición y que representa el color estimado, esta medición solamente representa una pequeña muestra de la tonalidad de la superficie total de la semilla, debido a que el equipo solo obtiene un valor puntual del color justo en el lugar donde incide el rayo de luz que se emite para obtener la medición, y para contender con esta limitante de medición en la práctica se toma la lectura en varios puntos de la superficie y a partir de los cuales se calcula un promedio (Figura 1.1).

No.	Date, Time	Población	L*	a*	b*
1	26/03/17 12:28	POB-02	73.33	1	10
2	26/03/17 12:28	POB-02	72.52	0.95	10.14
3	26/03/17 12:29	POB-02	71.42	0.97	8.52
4	26/03/17 12:29	POB-02	70.87	0.91	8.64
5	26/03/17 12:29	POB-02	72.22	0.87	9.87
6	26/03/17 12:29	POB-02	71.52	0.82	10.03
7	26/03/17 12:29	POB-02	73.44	0.74	11.06
8	26/03/17 12:29	POB-02	72.72	0.69	11.23
9	26/03/17 12:30	POB-02	73.1	0.69	10.73
10	26/03/17 12:30	POB-02	72.55	0.63	10.88



FIGURA 1.1: Medición de color. (A)Valores de medición, (B)Espectrofotómetro

Existen poblaciones nativas de frijol conformadas de granos de color uniforme, así como poblaciones con granos de frijol de dos colores y poblaciones con granos de tres o más colores, además los granos pueden tener pigmentos de otro color distinto al color predominante, existe gran variabilidad de color en las poblaciones nativas de frijol, la Figura 1.2 muestra las imágenes de ejemplos de poblaciones con distintos colores.

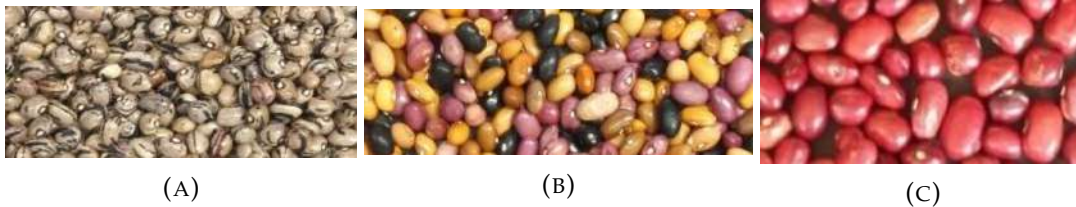


FIGURA 1.2: Variabilidad de color en las poblaciones nativas de frijol. (A)Población con granos de dos colores, (B)Población de granos de varios colores, (C)Población con granos de color uniforme

En la mayoría de los casos las poblaciones nativas de frijol se componen de semillas de diferentes colores, lo que hace inasequible el uso de esta técnica de medición de color con un espectrofotómetro, por el hecho de no poder considerar toda la distribución de tonos de color que tiene cada grano de frijol, la medición resulta subjetiva considerando el criterio del experto la cantidad de mediciones sobre el grano y los granos de la población.

Como información adicional, además de los granos de frijol ésta limitante también se presenta en la medición de frutos y hortalizas, ya que en un fruto u hortaliza el color no es homogéneo y aun cuando prevalezca un matiz este varía su tonalidad a través de la superficie total, la Figura 1.3 muestra una imagen de tomate y una imagen con semillas de frijol. En ambos casos se pueden apreciar las diferentes tonalidades existentes dentro de un mismo ejemplar, así como entre ejemplares.

Este mismo problema se presenta en los granos de frijol, en donde un mismo ejemplar de grano no contiene un tono único de color sino una distribución de tonos similares que le dan una coloración uniforme y que puede contener un degradado de coloración o pigmentos de otro color.

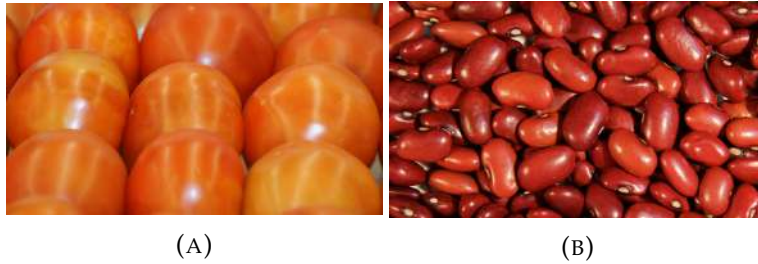


FIGURA 1.3: Variación en la tonalidad en frutos y semillas. (A)Tomate, (B)Frijol

Para contender con esta problemática algunos trabajos reportados en la literatura han utilizado sistemas de visión por computadora en los que utilizan una cámara CCD para la adquisición del color. Desafortunadamente, aun cuando este tipo de adquisición de imagen permite captar la superficie total a analizar, en la práctica, de manera similar al espectrofotómetro solo se han obtenido muestras puntuales o promedio de regiones de interés, lo que resume de una manera parcial e incompleta la distribución de color existente en el objeto analizar.

En este trabajo se propone un método que considera toda la distribución de tonos de color de los granos de las poblaciones nativas de frijol, para obtener una completa representación con base en la caracterización del color que permita la caracterización de poblaciones nativas de frijol.

A continuación, se describen algunos trabajos que han abordado la problemática de la clasificación de semillas de frijol, en ellos se resaltan las técnicas de análisis de color utilizadas, posteriormente se confrontará la técnica de medición puntual contra la metodología propuesta en este estudio.

1.1. Motivación

El problema que actualmente se presenta para la clasificación de poblaciones de frijol por color, es que para la evaluación del color del grano de frijol se registran los valores de una muestra puntual o promedio de color de una región de interés obtenidos con un espectrofotómetro, lo cual no resulta representativo y es limitado al no considerar todas las variantes que se pueden tener en la muestra real de población de frijol.

De los datos obtenidos de la medición, como complemento se realizan cálculos adicionales McGuire [24], para dar mayor interpretación a los valores de a^* y b^* , se incorporan el ángulo de matiz (h°) que resulta de la tangente inversa del valor en el canal a^* entre el valor del canal b^* y la cromaticidad que es una hipotenusa del triangulo formado por los puntos conformados por el valor del canal a^* y el valor del canal b^* , con lo que se determina la pureza del color Figura 1.4, las formulas para calcular estos valores son: $[C^* = (a^{*2} + b^{*2})^{1/2}]$, y $[h^\circ = \tan^{-1}(\frac{b^*}{a^*})]$, donde $a^* \neq 0$.

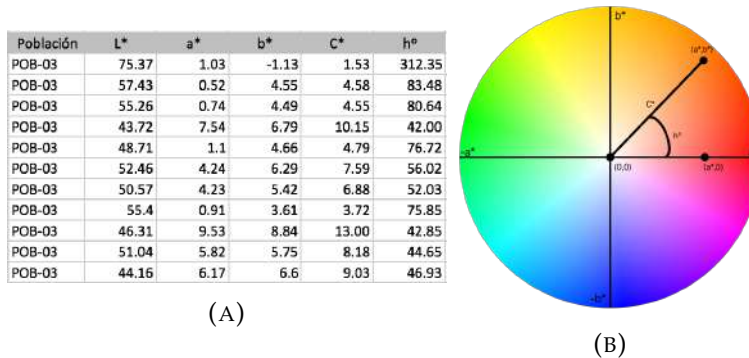


FIGURA 1.4: Medición de color. (A)Parámetros de color, (B)Modelo de color con ángulo de matiz y cromaticidad

Un promedio o conjunto de mediciones puntuales no cubren la gama de colores que puede presentar una semilla de frijol, la Figura 1.5 muestra como aun cuando la distribución de colores del grano es homogénea, esta no contiene un tono único ya que estos se presentan como una función de distribución de probabilidad, lo que se puede observar claramente en los histogramas tanto en la luminosidad como en los parámetros a^* y b^* .

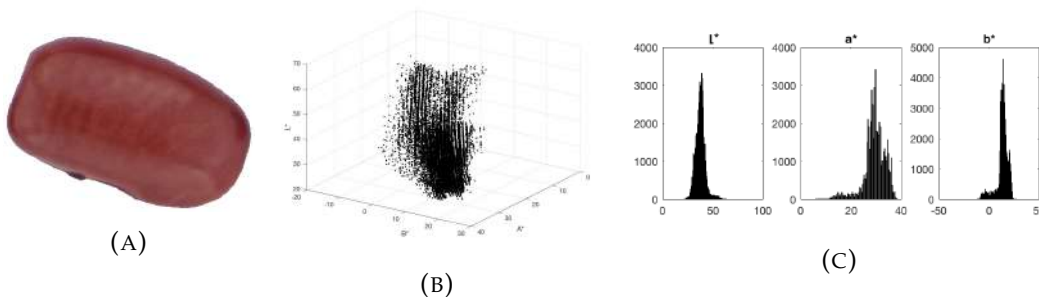


FIGURA 1.5: Ejemplo de distribución de color de grano de color negro; (A) Grano de frijol color negro, (B)Distribución de matices de color en el espacio CIE $L^*a^*b^*$, (C)Histogramas de cada componente de color L^* , a^* , b^*

La Figura 1.6 muestra que la distribución de color del grano no es homogénea, lo que se ve reflejado en la desviación estándar y adicionalmente se puede apreciar que no necesariamente tienen distribuciones unimodales. Este caso ejemplifica la problemática de reducir a una sola tonalidad el color representativo de un grano de frijol conformado por dos tonalidades.

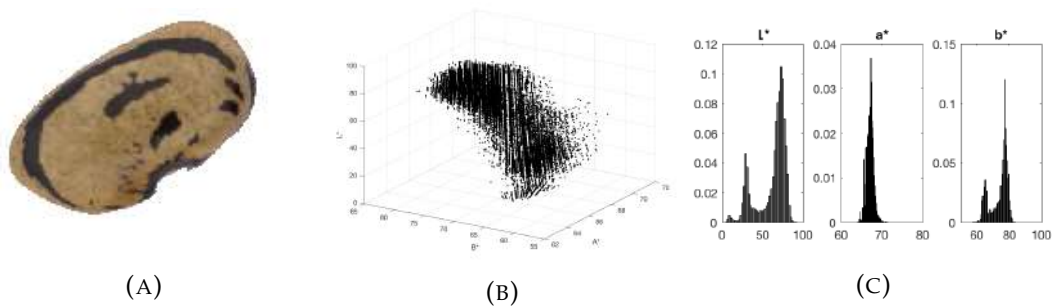


FIGURA 1.6: Ejemplo de distribución de color de grano de dos colores; (A) Granos de frijol con dos tonalidades, (B) Distribución de matices de color en el espacio CIE $L^*a^*b^*$, (C) Histogramas de cada componente de color L^* , a^* , b^*

De las poblaciones de granos de frijol analizadas en este estudio algunas contienen una gran diversidad de granos de distintos colores, las poblaciones con granos de color uniforme mantienen una amplia distribución de tonos de color similares, la Figura 1.7A muestra los granos de color blanco con similitud de tonos de color, existen otras tantas poblaciones que contienen una variedad de distintos granos de color, por lo tanto un promedio de color no aportaría una correcta caracterización de las poblaciones de frijol Figura 1.7B.



FIGURA 1.7: (A) población 02 y (B) población 33

Para la clasificación de las poblaciones nativas de frijol se requiere de un método que permita la utilización de toda la distribución de color que logre caracterizarlas, por lo cual los promedios o mediciones puntuales realizadas con un espectrofotómetro no son representativos para cubrir la necesidad.

La Figura 1.8 muestra la nube de distribución de tonos de color de los granos de la población de frijol 51, por lo que un conjunto de mediciones o un promedio de ellas no es muy representativo para la identificación. La nube de puntos de color de los 8 granos muestra claramente como los granos de color negro tiende a tener valores de luminosidad cercanos a cero y los granos de colores más claros están distribuidos en la zona de los valores más altos de luminosidad, el punto rojo muestra el promedio de color de los granos, lo que no resulta muy representativo.

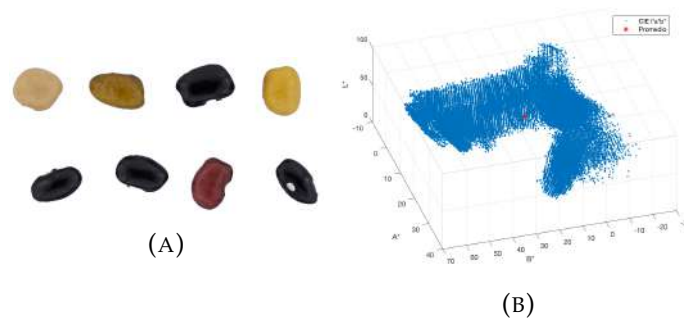


FIGURA 1.8: Ejemplo de distribución de color de grano de diversos colores; (A) Granos de la población 51, (B) Distribución de matices de los granos

En este trabajo se considera la distribución de color que conforman los 20 granos de las poblaciones, porque el color del grano de frijol que solo parece cubierto con un solo tono de color, contiene una distribución de tonos no homogéneos, considerando que un promedio de color no permitirá la correcta identificación de las poblaciones nativas de frijol, este trabajo ofrece un método que permita la correcta caracterización de las poblaciones utilizando toda la distribución de tonos de color de los granos para la identificación de acuerdo a la variedad local y la identificación por coloración de los granos, para ello se hace uso de histogramas y se propone representar como una distribución de probabilidad conjunta de los granos de frijol de cada población nativa de frijol.

1.2. Justificación

Dada la problemática actual para clasificar poblaciones nativas de frijol a través de las características colorimétricas, obtenidas mediante un equipo espectrofotómetro que registra mediciones puntuales de los granos que conforman cada población nativa de frijol, y los estudios previos reportan métodos de medición puntuales y que los sistemas de visión actuales reportan promedios de color, en el presente estudio se propone la representación del color total de los granos de frijol que conforman a cada población, mediante la distribución de su tonalidad en los espacios de color HSI y CIE L*a*b*.

Esta representación estará dada por histogramas en 2 y 3 dimensiones, que se utilizarán para ser evaluada mediante el algoritmo de clasificación K-Vecinos más cercanos, que permita comparar muestras considerando histogramas que representan la distribución de probabilidad conjunta característica de cada población, con la finalidad de lograr identificar las poblaciones nativas de frijol, y que además permitirá obtener información sobre poblaciones con similitud colorimétrica.

1.3. Hipótesis

Con base a la utilización de toda la distribución de color de los granos de frijol de cada población, es posible tener una mejor caracterización de las características colorimétricas y por lo tanto éste método permitirá una mejor identificación y clasificación de las 54 poblaciones nativas de frijol (*Phaseolus vulgaris* L.).

1.4. Objetivos

1.4.1. Objetivo General

- Identificar las propiedades colorimétricas de 54 poblaciones nativas de frijol mediante toda la distribución de color de los granos de cada población para obtener mejor representación y caracterización por color.

1.4.2. Objetivos Específicos

- Integrar histograma de frecuencias como un descriptor para representar toda la distribución de color
- Utilizar histogramas ocupando los espacios de color HSI y CIE L*a*b* para caracterizar el color de las poblaciones de frijol
- Aplicar uso de histogramas en el algoritmo de clasificación K-Vecinos más cercanos, para identificar poblaciones nativas de frijol

Capítulo 2

Antecedentes

La característica del color de un producto puede ser medido empleando visión por computadora, existe trabajos relativos a la clasificación de frutas y hortalizas de acuerdo a su textura, color, tamaño, forma, empleando los 3 tipos de sistemas de visión por computadora, como el sistema tradicional, el sistema de visión hiperespectral y el sistema de visión multiespectral[7]. Los anteriores métodos también han sido utilizados para la clasificación de distintos tipos de semillas en donde el proceso de clasificación manual es complejo y se realiza con base a distintos criterios tales como: textura, color y/o tamaño, el cual requiere de entrenamiento para una adecuada selección, son importantes para la decisión de compra de los consumidores.

Los sistemas de visión se han empleado en la clasificación de frutas y hortalizas exitosamente, evitando inconsistencias como tiempo, variabilidad, subjetividad, donde el proceso es tedioso, laborioso, costoso y fácilmente influenciado por el ambiente, surge la necesidad de desarrollar sistemas de inspección automática que reemplace la selección manual, los sistemas de visión proveen un medio para la realización de tareas automáticas y son una disciplina que se encarga del estudio y aplicación de métodos que posibiliten a una computadora entender el contenido de una imagen y extrayendo ciertas características importantes para el reconocimiento de patrones[7].

Los sistemas de visión por computadora se desarrollaron como una herramienta de inspección científica para la calidad y seguridad de alimentos y productos agrícolas [7], se han empleado exitosamente en la clasificación de granos y semillas. La identificación de granos de lentejas por medio de su tamaño y color, utilizando un escáner para la adquisición de imágenes[30], de igual manera los sistemas de visión han sido utilizados eficientemente en la clasificación de semillas de maíz considerando 13 tipos de defectos usando características de color y textura[19], y para la identificación de la pureza del maíz[32], también en la clasificación de semillas de colza[18] y clasificación de semillas de maleza para prevención de las malas hierbas[22].

Para la clasificación manual de semillas de frijol se han realizado trabajos relativos a la medición del color con un espectrofotómetro para sólidos, realizando varias mediciones puntuales de color de forma manual[3]. Este método presenta los siguientes problemas, el color del grano no es homogéneo como consecuencia las lecturas puntuales son distintas, se tiene la subjetividad de medir en varios puntos en un mismo frijol y es un proceso complicado cuando la lectura es para varios granos de distintos colores de una población de frijol.

De los trabajos realizados utilizando sistemas de visión por computadora, se tiene el reporte que utilizó el análisis de imagen digital para la medición de las características morfológicas de las semillas de frijol en cinco diferentes cultivos (Akhtar, Ks21189, Ks31169, G11867 y khomein) Figura

2.1, considera las propiedades morfológicas como el área, perímetro, área convexa, diámetro, entre otras características para mostrar diferencias en las semillas de los cultivos, sus resultados indican diferencias significativas entre los cultivos en características morfológicas que fueron medidos[26]. Este estudio no considera las propiedades colorimétricas de los granos para caracterización y no utiliza un algoritmo de aprendizaje automático.



FIGURA 2.1: 5 tipos de frijol

En el estudio reportado por Gianfranco Venora y colaboradores emplean un método de clasificación para 15 variedades de frijol criollo Figura 2.2, con el objetivo de identificar a cada variedad y discriminar el año de cultivo para estimar la posible influencia del entorno de crecimiento con relación a los parámetros morfológicos de la semilla, y descubrir las posibles diferencias en las variedades locales de las diferentes regiones de cultivo, de las 15 variedades de frijol una variedad se consideró para evaluar con los granos de dos regiones de cultivo y dos variedades para evaluar con granos de diferentes años de cosecha. En éste trabajo se utilizó un escáner (ScanMaker 9800XL) para la adquisición de imágenes, obteniendo las características físicas de tamaño, forma, color y textura de cada grano, además, obtienen la media y desviación estándar de cada componente de color del modelo RGB (los canales Red, Green, Blue) y modelo HLS (Hue, Ligth, Saturation) de toda la semilla, con un total de 26 variables cualitativas morfológicas y colorimétricas realizan la implementación del método de análisis discriminante lineal.

El desempeño de las clasificaciones fue el siguiente: el primero, para la identificación de 15 variedades autóctonas, 99.00% de clasificación del conjunto de entrenamiento y 98.20% de precisión en la clasificación del conjunto de pruebas; El segundo, para identificar dos años de cosecha, el 99.00% de precisión para los ejemplos del conjunto de entrenamiento y el 100% de clasificación para granos de la misma variedad; El tercero, para identificar tres años de cosecha, 100% de clasificación para el conjunto de entrenamiento y 98.82% para la clasificación de los grano de la misma variedad; El cuarto, para la identificación de las dos regiones de cultivo diferentes, 94.5% de precisión en la clasificación de los ejemplos del conjunto de entrenamiento y 82.40% de clasificación en granos de la misma variedad. [31]. Este estudio utiliza el promedio de color en los espacios de color RGB y HSL, aunque lo complementa con variables morfológicas y logran precisiones de clasificación altas, éste método no cubre la utilización de toda la distribución de tonos de color del grano de frijol.

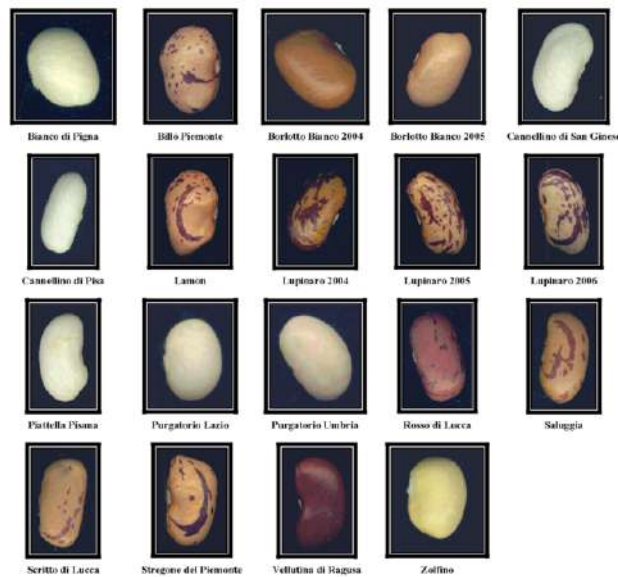


FIGURA 2.2: 15 variedades de frijol y 4 con distinto años de cosecha

Por otra parte A. Nasirahmadi y N. Behroozi-Khazaei para la identificación de diez variedades de granos de frijol Figura 2.3 de las variedades Khomein1, KS21108, Khomein2, Sarab1, Khomein3, KS21409, Akhtar2, en este trabajo utilizan una cámara con la que adquieren imágenes de 576 x 720 píxeles, 12 características de color se extrajeron de las imágenes de frijol. Se trata de componentes de color RGB que representa los canales (Red, Green, Blue) y el modelo HSI que significa los canales de (Hue, Saturation, Intensity) obteniendo el promedio y desviación estándar de cada canal, los dos componentes de color para el color de toda la semilla de frijol y de las manchas de color del grano de frijol, implementan redes neuronales artificiales para la identificación de las variedades de frijol y obtienen resultados de clasificación con una sensibilidad general del 100 %, 97,33 %, 96 % y especificidad de 100 %, 97,9 % y 97,1 % para entrenamiento, validación y prueba, respectivamente [25]. Para caracterización de las 10 variedades hacen uso de 12 características del color del grano, 6 características conformadas por la media de cada componente de color de los espacios de color RGB y HSI obtenidos de toda la semilla y 6 características de las manchas de la semilla donde cada una es también un promedio de cada componente de color, para evitar el problema que se tienen con el promedio de cada espacio de color integran un elemento distintivo como lo es la mancha de color del grano con el que identifican variedades con buenas precisiones, solo que este enfoque no utiliza toda la distribución de tonos del color para caracterizar y a pesar de tener buenas precisiones tienen solo promedios para pocas variedades.

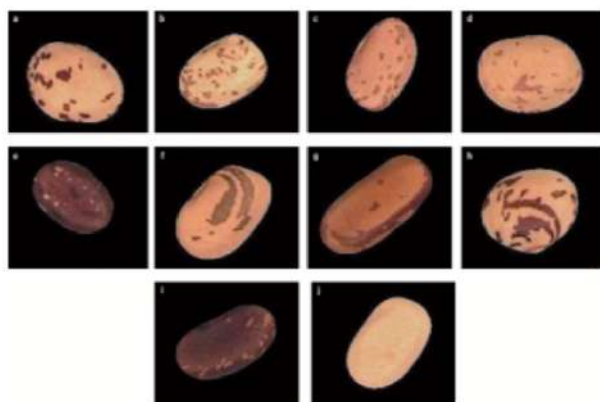


FIGURA 2.3: 15 variedades de frijol y 4 con distinto años de cosecha

En el reporte de Sidnei Alves de Araujo y colaboradores reportan clasificación de tres variedades de frijol, carioca, mulato y negro Figura 2.4, indicó que el frijol carioca es la variedad de frijol más consumida en Brasil dominando el 70% del mercado, el frijol negro el 20% y el resto de las variedades el 10%. Por otra parte, el frijol negro y el frijol mulato son los dos granos menos predominantes que se encuentran con mayor frecuencia en lotes de frijol carioca, el problema es que un lote de frijol sólo puede ser etiquetado como Carioca si contiene no más del 3% de granos negros y 10% de granos de mulato. Este estudio emplea un sistema de visión que utiliza una cámara web para obtener imágenes de los granos, de la imagen mapean los pixeles de los granos con color negro como primer plano y los valores de los pixeles del fondo de la imagen con color blanco, cuando el sistema no puede decidir si el pixel es de primer plano o de fondo, se asigna al pixel un tono de gris, para realizar el mapeo toman algunos colores de las diferentes variedades de frijoles como datos de entrenamiento, y algunos pixeles del fondo son considerados como datos de entrenamiento negativos y dado un nuevo pixel lo asigna a la clase del pixel con el color más similar utilizando el algoritmo de aprendizaje K vecinos cercanos. Posteriormente para la segmentación de las semillas utilizan la técnica de correlación basada en granulometría, que consiste en una plantilla de todas las posibles formas de un frijol. Después de la localización de los granos que pueden ser únicamente clasificados como carioca, mulato o negro, utilizan el color predominante de cada grano en el espacio de color RGB. El frijol negro puede ser fácilmente reconocido por que su color predominante que es muy diferente al color del frijol carioca y mulato, el color predominante del frijol carioca es beige pero algunos granos tienen café marrón y el color predominante del frijol mulato es marrón rojizo pero algunos granos pueden tener marrón o beige, el color café marrón del frijol carioca es muy similar al color predominante del frijol mulato y algunas veces el color marrón ocupa una considerable área del frijol, algunas partes del frijol mulato pueden tener color muy similar al color del frijol carioca.

Como se tiene granos de dos colores se utiliza el algoritmo K-Means para la separación de los colores, donde más del 70% de los pixeles se considera color dominante del grano, después de la extracción del color predominante del grano para la clasificación nuevamente utilizan al algoritmo K-vecinos cercanos, obtienen resultados de clasificación del 99.88% de precisión para la identificación de las tres variedades[5]. La precisiones de este trabajo son altas considerando que solo identifican tres distintas variedades de frijol utilizando el color más predominante en el grano, a pesar que utilizan un conjunto de pixeles para identificar a cada tipo de grano no es suficiente para la caracterización de poblaciones nativas de frijol considerando que se tiene

poblaciones de granos muy similares, no resulta apropiado usar solo el color predominante, y aun que para la integración de más variedades recomiendan hacer uso de la textura no resulta adecuado para cubrir la necesidad de utilizar toda la distribución de tonos de color de los granos para caracterizar a las poblaciones.



FIGURA 2.4: 3 variedades de frijol

Bitjoka Laurent y colaboradores emplean un método para la identificación de 16 tipos de frijol usando matriz de filtro de color (CFA), para la adquisición de las imágenes se utilizó una cámara digital CCD con su máxima resolución 3664x2742 pixeles, donde los granos fueron dispersos sobre la superficie, este trabajo usó el software ImageJ para determinar el área y perímetro de la superficie de la semilla que se utilizará para determinar la compacidad, la compacidad refleja la naturaleza compacta de las semillas de frijol. La compacidad se define por: $C = \frac{2}{Perimeter} \sqrt{\pi (Surface\ area)}$, los resultados se comparan con los valores del método de elementos finitos que son determinados usando medidas físicas, los resultados muestran que los granos de frijol variedad DOR-701 son más compactos en comparación con las otras variedades de granos de frijol analizados. El análisis de los resultados obtenidos muestra que el coeficiente de variación se sitúa en torno al 5% [21]. El problema es que solo se muestra una comparativa de formas usando análisis digital de imágenes y el método de elementos finitos, solo hace uso de las propiedades morfológicas y de los granos no usa las propiedades colorimétricas y tampoco emplea un algoritmo de clasificación, por lo que no resulta adecuado para cubrir la necesidad de hacer uso de la coloración de los granos para caracterizar e identificar poblaciones nativas de frijol.

















DOR-701		GLP-190(F.V.)	
MERINGUE (PH-201)		MAC-16	
MAC-33		MAC-18	
SENEGALAIS		MAC-55	
CIM 9331-1		MPOLO	
ECA PAN 019		TY 3396-12	
ECA PAN 021		Vtt 926 5-7	
GLP-(F.B.)		ZEBRA	

FIGURA 2.5: 16 tipos de frijol

El método de identificación de 15 tipos de frijol criollo empleado por Gianfranco Venora [31], es mejorado por Marisol Lo Bianco y colaboradores en 2015 para la identificación de 67 tipos de semillas de frijol de Italia, añadiendo algoritmos que permiten calcular los Descriptores Elípticos de Fourier(EFDs) para cada semilla analizada, con lo que obtienen 78 variables cuantitativas, este método permite definir el contorno de la semilla como un conjunto de números que corresponden a la posición de píxeles del límite de la semilla, además se mejoró la macro que incluía algoritmos capaces de calcular 11 Descriptores de Haralick con los valores de desviación estándar relativos para cada semilla, los descriptores de Haralick se miden matemáticamente en cada semilla como: La textura de la superficie y todas las características morfológicas. Los resultados por color para blanco, monocolor y una capa de semilla bicolor o manchada aplicando el modelo de discriminación, los porcentajes de identificación correcta oscilaron entre el 98.2% en granos monocolor y el 100.0% en granos blancos, con un total de 99.1%, lo que confirma la importancia de las características del color para la discriminación de los granos. Posteriormente, se desarrolló una clasificación para cada una de las tres categorías principales de color de la capa de semilla de frijol, considerando únicamente el nombre de la raza local como variable de agrupación. En cuanto a las 13 variedades locales blancas, el porcentaje global de la clasificación correcta fue del 96.0%, el más bajo se registró para Cannellino di Pisa con 85.8% y el más alto para Triverde con

100.0 %; Mientras que, las ocho semillas de frijol monocolor de haba alcanzaron una clasificación correcta global del 99.7%, registrando para Giallo, Moitano, Tabacchino y Vellutina di Ragusa la tasa de identificación perfecta del 100 % [23]. En este trabajo para dar mayor clasificación integran nuevos algoritmos para la utilización de características morfológicas y de textura, en cuanto al color siguen utilizando un promedio y este enfoque es completamente diferente al que se desea para la clasificación de variedades utilizando toda la distribución de tonos de color de los granos.



FIGURA 2.6: 67 tipos de frijol

El análisis de Bitjoka Laurent y colaboradores utiliza sistema de visión por computadora para determinar el nivel de dificultad de cocción de la alubia roja, mediante la utilización de la propiedad físicas como el color, para generar histogramas de los canales separados R,G,B y el canal L^* de luminosidad de CIE $L^*a^*b^*$, de cada histograma obtienen los máximos y mínimos valores de niveles de gris de cada componente, para verificar la variación en el color del frijol y determinar que los granos experimentan cambios de color durante el almacenamiento, lo cual se relaciona con el fenómeno de difícil cocción, para este trabajo solo utilizan comparativa de histogramas de cada componente de color para evaluar la relación del almacenamiento con la cocción [8], a pesar de que este análisis hace comparativa de histogramas sobre granos almacenados, solo emplea el espacio de color RGB y del espacio de color CIE $L^*a^*b^*$ el canal de luminosidad (L^*), no utiliza algún algoritmo de aprendizaje automático y está destinado solo a la verificación del cambio de color de almacenado.

El cuadro 2.1 muestra comparativa de trabajos relacionados a la clasificación de granos de frijol y las características significativas de cada uno, incluye materiales de estudio, medio de adquisición de imágenes, algoritmos de segmentación, las características externas de los granos que utilizaron y el algoritmo de clasificación empleado para la identificación de los tipos de granos.

Los recientes trabajos integran las características físicas, morfológicas y colorimétricas, con respecto al color registran un promedio de todo o parte de una región de interés del frijol y para la clasificación hacen uso de algoritmos como análisis discriminante lineal, redes neuronales o k-vecinos cercanos. El problema es que un promedio de color o medición puntual de color de granos de frijol no es representativo, la finalidad de éste trabajo es considerar toda la distribución de tonos de color que conforman a los granos de frijol de cada población nativa de frijol, para tener una mejor caracterización de 54 poblaciones nativas de frijol.

CUADRO 2.1: Comparativa de trabajos relacionados con la identificación de variedades de frijol, se despliega las características utilizadas para la clasificación.

Autores	Materiales	Dispositivo de adquisición	Algoritmo de segmentación	Características utilizadas	Algoritmo de clasificación	Observaciones
Laurent Bitjoka y colaboradores, 2010	Frijol rojo	Cámara digital CCD Fujifilm FinePix F480	Umbralización	pixeles de color del grano en RGB, histogramas nivel de gris para R, G y B	-	Solo comparativa de histogramas para determinar el nivel cocción de uan variedad de frijol rojo
Gianfranco Venora, 2009	15 tipos de frijol	Escáner	Filtro adaptativo	Características físicas de tamaño, forma, del color obtienen promedio y desviación estándar de color, en el espacio RGB y HLS, para toda la semilla, manchas de color de la semilla	Método de Análisis de Discriminantes lineal	Enfoque utiliza promedio de color y se apoya de características morfológicas
Seyed Saeid Modarres Najafabadi, Leila Farahani 2012	5 tipos diferentes de frijol	Cámara digital (Canon G7)	Detector de bordes con Laplaciana de la gaussiana	Características morfológicas (área, perímetro, área convexa, diámetro)	-	Análisis de variación de características morfológicas
A. Nasirahmadi and N. Behroozi-Khazaei 2013	10 tipos diferentes de frijol	Cámara digital Samsung	Umbralización a los canales R y B	12 variables de color de las semillas utilizando promedio de los espacios de color RGB y HSI	Redes neuronales artificiales	Este enfoque utiliza un promedio de color
Sidnei Alves de Aratújo, Jorge Henrique Pessota, Hae Yong Kim 2015	3 tipos diferentes de frijol	Cámara Microsoft LifeCam HD-5000	Correlación basada en granulometría	Etiquetan conjuntos de pixeles de los colores predominantes de los granos	K vecinos más cercanos	Utilizan un conjunto de pixeles para identificar y sugieren hacer uso de la característica de textura para mejorar enfoque, enfatizan su contribución en su método de segmentación
Laurent Bitjoka y colaboradores, 2015	16 tipos diferentes	cámara digital CCD FUJIFILM Finepix HS10)	no especificado, ImageJ software®	Características morfológicas de los granos	-	Este trabajo solo hace una comparativa de mediciones morfológicas
Marisol Lo Bianco, Oscar grillo y colaboradores 2015	67 tipos de frijol	Escáner	Filtro adaptativo	Características físicas de tamaño, forma, del color obtienen promedio y desviación estándar de color, en el espacio RGB y HLS, para toda la semilla, manchas de color de la semilla, mejoran enfoque con Descriptores Elipticos de Fourier y Descriptores de Haralick	Método de Análisis de Discriminantes lineal	Retoma trabajo y tiene que aplicar mejoras para poder integrar más variedades

Capítulo 3

Metodología

La metodología de este aporte propone caracterizar poblaciones nativas de frijol mediante sus propiedades colorimétricas, como una alternativa viable comparada con la evaluación del color ofrecida por un espectrofotómetro, y para lograr el objetivo de este trabajo se implementó la siguiente metodología, que consiste en la adquisición, segmentación, extracción de características e implementación de aprendizaje automático como lo muestra la Figura 3.1.

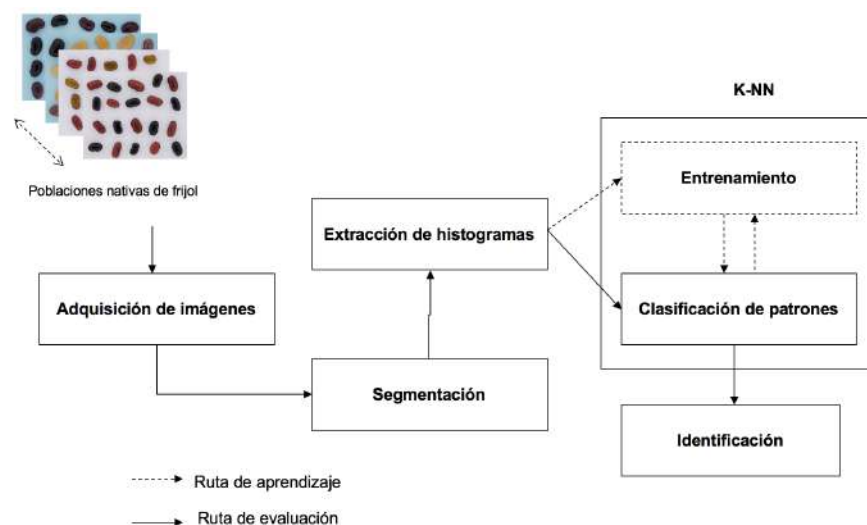


FIGURA 3.1: Metodología propuesta para la caracterización de poblaciones nativas de frijol

En la adquisición de imágenes se implementó un prototipo de iluminación como un entorno controlado de iluminación para la adecuada adquisición de las imágenes, esta contribución a la metodología fue crucial para la correcta caracterización de las poblaciones nativas de frijol, además con un fondo bien definido sin sombras el algoritmo de segmentación utilizado en este trabajo resultó muy útil para quitar el fondo a los granos de frijol de cada población. El algoritmo K vecinos más cercanos resultó de gran utilidad para la clasificación e identificación de las poblaciones de acuerdo a la variedad local como a su clasificación por color de los granos, superando los porcentajes de clasificación obtenidos con el espectrofotómetro.

3.1. Materiales y métodos

3.1.1. Poblaciones de granos de frijol

Las poblaciones de granos de frijol empleadas en este trabajo están conformadas por veinte granos cada muestra y proceden de distintas regiones del estado de Oaxaca. Las poblaciones fueron proporcionadas por El Dr. José Luis Chávez Servia del CIIDIR-Oaxaca del Instituto Politécnico Nacional al Instituto de Ciencias Básicas de la Universidad Veracruzana para la realización de este trabajo. El apéndice A muestra los datos de las georreferencias de cada lugar donde fueron colectadas las poblaciones y que son variedades locales de lugares donde el cultivo de frijol es la fuente principal de alimento. Dichas poblaciones están conformadas de colores uniformes como rojo, blanco, negro, amarillo, otras poblaciones de frijol de la combinación de granos de dos colores, por ejemplo, rojo y negro, blanco y rojo, también existen poblaciones con granos de diversos colores, adicionalmente se tiene granos con dos colores o que contienen pigmentos de otro color consultar apéndice B.

3.1.2. Software

Para la realización de este trabajo y de la creación de los algoritmos se usó el software Matlab®2012.

3.2. Modelos de color

3.2.1. Espacio de color

Un espacio de color es un modelo que puede ser usado para representar colores, por color podemos referirnos a los colores que el sistema visual humano puede percibir y varios colores pueden ser obtenidos por todas las posibles combinaciones de todos los colores del espectro visible. El intervalo de colores posibles que puede ser representado por un sistema particular se llama gama. Un espacio de color en su forma más simple, puede ser visto como la combinación de tres colores primarios rojo, verde, azul. Añadir cualquiera de estos colores entre sí puede producir una amplia gama de colores, ésta es la base del modelo de color RGB [1].

Los modelos de color son útiles para tener una estandarización de color, esto evita problemas subjetivos en la definición y caracterización de colores, el espacio de color RGB es una representación útil para la visualización de imágenes de color en los dispositivos como las pantallas, escáneres, cámaras, entre otros, las impresoras manejan colores del espacio de color CMYK, y así como estas representaciones de color tienen utilidad, existen otras con distintos propósitos, como HSI y CIE L*a*b* que permiten tener una representación más acorde a la forma en como los humanos interpretan los colores separando la luminancia de la cromaticidad para fines de procesamiento digital de las imágenes.[14].

3.3. El modelo de color RGB

En el modelo RGB, cada color aparece como un componente espectral de rojo, verde y azul, este modelo está basado en sistema de coordenadas cartesiano, la Figura 3.2 muestra el cubo en el cual los valores de R, G, y B están en tres esquinas, y cian, magenta y amarillo en otras tres esquinas,

el negro se ubica en el origen y el blanco en la esquina más alejada del origen, en este modelo la escala de grises se extiende de negro a blanco a través de la línea que une a estos dos puntos. Los diferentes colores en este modelo son puntos dentro del cubo y están definidos por vectores que se extienden desde el origen. Las imágenes representadas en el modelo de color RGB consisten de tres componentes, uno para cada color primario, los valores de RGB que conforman un píxel de color es una tripleta de valores (R,G,B)[14]

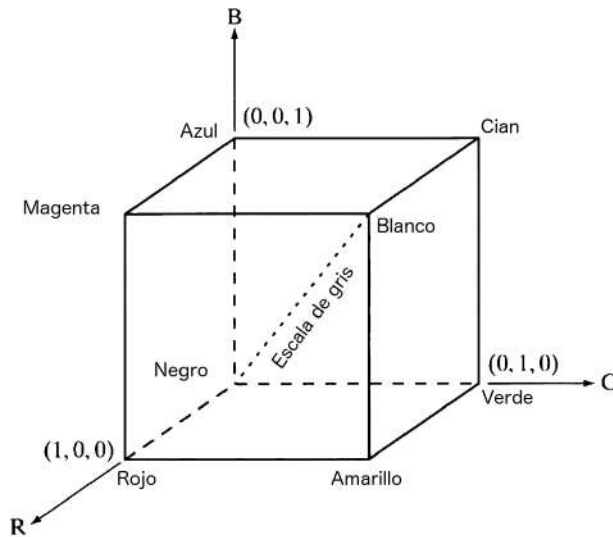


FIGURA 3.2: Espacio de color RGB

Este espacio de color está correlacionado, la suma del valor de cada canal produce el color y éste espacio de color no se utiliza como descriptor de color, porque es para mostrar imágenes en monitores o cámaras y para este trabajo se ocupará los espacios de color HSI y CIE $L^*a^*b^*$ como descriptores de color por estar más diseñados en la forma de como el ser humano interpreta los colores.

3.4. El espacio de color HSI

El espacio de color HSI es una herramienta ideal para el desarrollo de algoritmos de procesamiento de imágenes, basado en la descripción del color siendo natural e intuitiva para los humanos, este espacio de color compuesto de tres canales mantiene la separación del matiz, saturación e intensidad, trabajar con esta representación de color permite lograr un modelo deseable para la interpretación del color [14].

El componente de color hue es medido en grados y gira en sentido contrario a las manecillas del reloj para posicionarse sobre los colores principales. El componente saturación es la distancia del eje origen al punto de saturación y representa la pureza del color y el componente de intensidad que va de negro a blanco representa la intensidad luminosa, la Figura 3.3 contiene un esquema gráfico en forma cónica del espacio de color HSI.

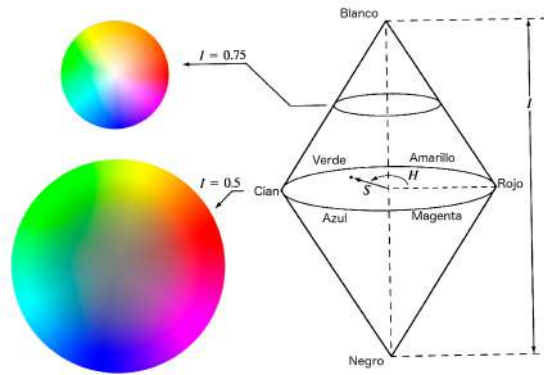


FIGURA 3.3: Espacio de color HSI

3.4.1. Conversión de espacios de color RGB a HSI

Para hacer uso del espacio de color HSI, es necesaria la conversión de una imagen RGB pixel por pixel con la siguiente ecuación, con la finalidad de disponer de la información colorimétrica teniendo separada la cromaticidad de la luminosidad, para la conversión es necesaria la normalización de los valores de RGB al rango [0-1].

$$H = \begin{cases} \theta & \text{si } B \leq G \\ 360 - \theta & \text{si } B > G \end{cases} \quad (3.1)$$

con:

$$\theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R - G) + (R - B)]}{[(R - G)^2 + (R - B)(G - B)]^{\frac{1}{2}}} \right\}, \quad (3.2)$$

θ : No está definida para valores de gris cuando $R=G=B$

El componente de saturación está dado por:

$$S = 1 - \frac{3}{(R + G + B)} [\min(R, G, B)], \quad (3.3)$$

S : Está definida cuando $(R,G,B)=(0,0,0)$

El componente de intensidad está dado por:

$$I = \frac{1}{3} (R + G + B) \quad (3.4)$$

3.5. El espacio de color CIE L*a*b*

CIELAB es un sistema adaptado por CIE en 1976 como modelo que mejor muestra un espaciado de color uniforme en sus valores. CIELAB es un sistema de color basado en un sistema anterior de Richard Hunter llamado L, a, b. El color se correlaciona con los descubrimientos a mediados de los años setenta, entre el nervio óptico y el cerebro, los estímulos de color de la retina se traducen en distinciones entre luz y oscuridad, rojo y verde y azul y amarillo. CIELAB indica tres valores con tres ejes: L*, a*, y b*.

La Figura 3.4 muestra el espacio de color, el eje vertical central representa la luminosidad (L*) con valores de 0 (negro) a 100 (blanco), los ejes de color van de rojo a verde y de azul a amarillo, en cada eje el valor va de positivo a negativo. con el eje de a* el valor positivo indica cantidad de rojo mientras que los valores negativos indican cantidad de verde, en el eje de b* el amarillo es positivo y el azul es negativo, para ambos ejes el valor cero es gris neutro [17].

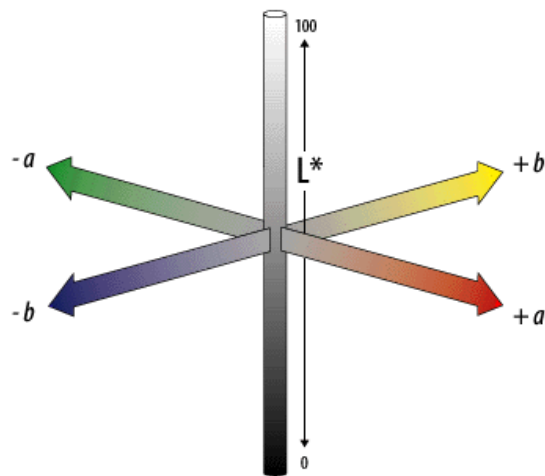


FIGURA 3.4: Espacio de color CIE L*a*b*

El espacio de color CIE L*a*b*, es un espacio ampliamente utilizado en la medición manual de sólidos y se ha utilizado para dar referencia colorimétrica en trabajos químicos incluido el frijol. Para este trabajo se consideró la utilización de este espacio porque de manera similar al espacio de color HSI se tiene la separación de cromaticidad y de la luminosidad, lo hace ideal para la representación del color.

3.5.1. Conversión del espacio de color RGB a CIE L*a*b*

Para la utilización del espacio de color CIE L*a*b* se hace uso de formulas de conversión, donde Y representa la luminosidad mientras X y Z son los componentes de color, para una conversión de RGB al espacio XYZ se utiliza la siguiente ecuación.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119194 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3.5)$$

Los valores para White Point D65 está dado por:

$$\begin{bmatrix} X_n \\ Y_n \\ Z_n \end{bmatrix} = \begin{bmatrix} 95.047 \\ 100 \\ 108.883 \end{bmatrix} \quad (3.6)$$

Para cada componente de color:

$$L^* = \begin{cases} 116 * \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 & \text{si } \left(\frac{Y}{Y_n}\right) > 0.008856 \\ 903.3 * \left(\frac{Y}{Y_n}\right) & \text{caso contrario} \end{cases}$$

$$a^* = 500 \left[\left(\frac{X}{X_n}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} \right]$$

$$b^* = 200 \left[\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n}\right)^{\frac{1}{3}} \right]$$

3.6. Adquisición de imágenes

Para la adquisición de imágenes se utilizó una cámara fotográfica con ajustes de configuración manual, además se implementó un entorno de iluminación uniforme para tener la iluminación ideal que permitiera visualizar el color de los granos, reduciendo el brillo de los granos y permitiendo a la cámara fotográfica capturar toda la gama de colores de los granos de frijol, los puntos siguientes describen las partes del entorno de iluminación y el procedimiento que permitió lograr la adquisición de las imágenes de este trabajo para la evaluación del color en los granos.

3.6.1. Prototipo de iluminación

El modelo de iluminación desarrollado en este trabajo tiene la finalidad de ofrecer un entorno iluminación constante y difuminada que sirve para reducir las sombras de los granos y reducir el brillo que tienen ciertas poblaciones nativas de frijol. El prototipo de iluminación fue elaborado a base de cartón con dimensiones de 30x30 cm en la base y 45 cm de altura, con luminarias en cada una de las esquinas para una adecuada dispersión de luz, las fuentes luminosas utilizadas por el prototipo son de 40 watts de luz cálida, con una conexión eléctrica en paralelo, en el interior del prototipo una caja forrada en tela de color blanco para la dispersión de la iluminación, reducción de sombras y de brillo en los granos. La Figura 3.5 muestra el esquema de prototipo.

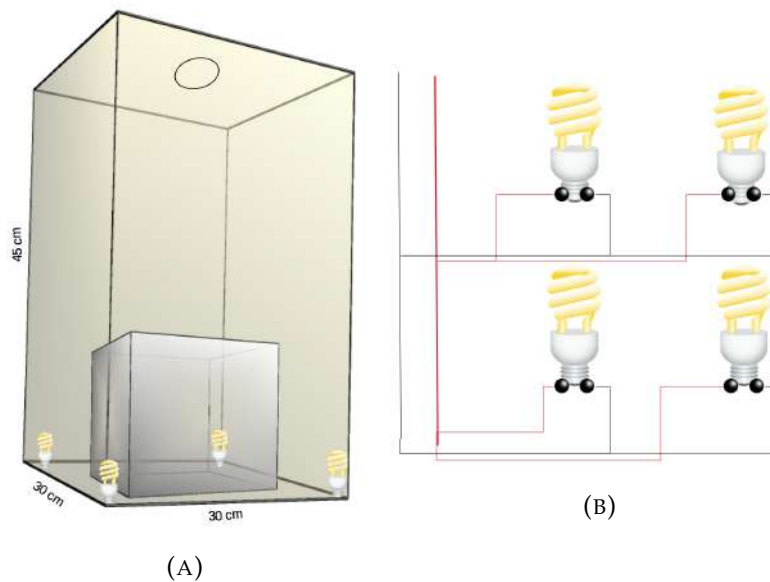


FIGURA 3.5: Esquema prototipo de iluminación. (A) Dimensiones del prototipo, (B) Conexión paralela eléctrica

La Figura 3.6b muestra el interior del prototipo, y de manera necesaria las fuentes luminosas fueron colocadas en cada una de las esquinas inferiores, con el propósito de tener el entorno completamente iluminado, ya que la intensidad de iluminación influye sobre la superficie de los granos de tal manera que con una adecuada iluminación se puede visualizar hasta los pigmentos pequeños de color sobre la superficie de los granos, en contra parte si la iluminación es deficiente se tendrán adquisiciones de imágenes con granos oscuros sin captar en la imagen esos detalles de color y lo importante para este trabajo es considerar toda la gama de colores en los granos que ayude a la correcta identificación de las poblaciones. La tapa del prototipo se forro con papel negro para evitar el reflejo de la luz en la parte alta, además contiene un orificio al centro que permite introducir la lente de la cámara.

Para la medición del entorno de luz se ocupó el exposímetro SEKONIC MULTIMETER L-408, en modo de ambiente de luz con ISO 100 se obtuvo el valor de exposición de 10 en el interior del prototipo, que de acuerdo a la tabla de conversión se tiene 3600 Lx como nivel de iluminación.

3.6.2. Caja de luz para fotografía

La caja de luz se utiliza en fotografía para la reducción de brillo en objetos de cristal y contiene el principio de difuminación de luz al evitar luz directa mediante un material translucido, para este trabajo se empleó como complemento del prototipo de iluminación, con la finalidad de reducir el brillo en algunas poblaciones de granos de frijol, que al recibir directamente la luz tiene un reflejo con forma de aro de luz sobre su superficie, y para contender con este problema la caja de luz reduce el brillo y evita que la luz incida directamente sobre los granos de frijol dejando a disposición la gama de colores que presentan los granos.

La caja de luz es un cubo forrado de tela blanca en los laterales, con dimensiones 19.5 cm x 19.5 cm x 19.5 cm, la Figura 3.6 muestra el ejemplo de caja de luz utilizada en este trabajo.



FIGURA 3.6: Prototipo de iluminación. (A)Caja de luz, (B)Prototipo de iluminación

Como parte de este trabajo se requiere hacer uso de las características colorimétricas, es necesario capturar hasta el más mínimo pigmento de color de la muestra, por lo que la iluminación debe ser constante y suficiente. La Figura 3.7 muestra tres imágenes para comparativa visual de adquisición. La imagen adquirida de los granos con escáner tiene el inconveniente de sombras, reflejos y manchas de luz provocadas por la luz directa recibida de la lámpara del escáner al momento de la digitalización, la imagen adquirida con prototipo de iluminación sin caja de luz las muestras reflejan la luz y este tipo de adquisición no resulta adecuado por la sección donde el brillo es prominente en el grano, y la imagen adquirida con prototipo de iluminación complementado con caja de luz son adecuadas para este trabajo y con difuminación de luz hay reducción de brillo y sombras.

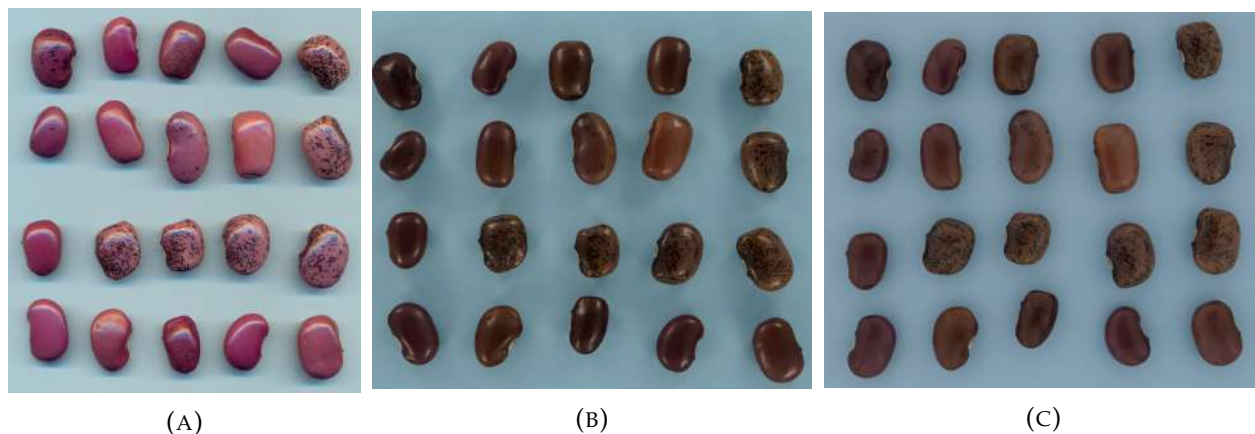


FIGURA 3.7: Granos de frijol de la población 22. (A) Imagen adquirida con escáner, (B) Imagen adquirida sin caja de luz, (C) Imagen adquirida con caja de luz

3.6.3. Cámara fotográfica

Una pieza importante dentro de los sistemas de visión por computadora es el dispositivo para la adquisición de las imágenes y en este trabajo se ocupó la cámara digital SONY modelo Ilce 3500 de 20MP, con configuración de disparo mostrada en el cuadro 3.1.

La configuración de la sensibilidad ISO debe ser de 100 para obtener imágenes nítidas, evitando el problema de sensibilidad, en el que a mayor sensibilidad el ruido en una imagen aumenta. El tiempo de exposición se consideró de 1/50 y la abertura de 5.6 adecuado para una imagen

CUADRO 3.1: Configuración de disparo de la cámara fotográfica.

Modo	Tiempo de exposición	Abertura
Manual	1/50	5.6
Zoom óptico	Enfoque	ISO
50mm	manual	100
Balance blanco	Área de enfoque	Modo de medición
incandescente	múltiple	Múltiple

no tan oscura ni tan clara. Por el tipo de lente de la cámara que no ofrece tomas fotográficas a corta distancia, se usó el zoom óptico de 50 mm para obtener imágenes más claras y definidas, se realizaron pruebas y resultó ser el zoom que mejor resultado, las otras opciones de zoom no daban el acercamiento necesario a la misma distancia, el balance de color en la cámara se indicó incandescente como especificación al tipo de fuente luminosa que se ocupó en el interior del prototipo de iluminación. Se usó la configuración manual porque la cámara permite lograr el mismo disparo de adquisición para todas las imágenes de las poblaciones, es importante desactivar la opción que corrige el brillo y contraste automáticamente, porque a pesar de que la iluminación sea constante las distintas imágenes pueden variar mostrando fondos de distinto brillo y contraste, y la finalidad es lograr imágenes que muestren la misma intensidad de iluminación, para evitar sesgos.

3.6.4. Proceso de adquisición de imágenes

El proceso para la adquisición de las imágenes consistió en colocar dentro de la caja de luz la muestra de 20 granos de frijol de forma separada para evitar que los granos se hicieran sombra y quedaran pequeños espacios que se pudieran formar entre el conjunto de semillas y que al momento de aplicar el algoritmo de segmentación no se considere la eliminación de esos espacios, para contrastar los granos de frijol de algunas poblaciones se ocuparon tres distintos colores de fondo que son el: blanco para las poblaciones de color negro y rojo y combinación de ambas, fondo negro para las poblaciones de color blanco y por último el color azul para las poblaciones de distintos colores. Posteriormente se realizó una toma fotográfica desde la parte superior central del prototipo de iluminación. El mismo procedimiento se realizó para las 54 poblaciones nativas de frijol, La Figura 3.8 muestra parte de la adquisición total de las poblaciones de frijol.



FIGURA 3.8: Pequeño conjunto de poblaciones de frijol

Las imágenes se adquirieron con una resolución de 5456 x 3632 píxeles en formato JPG, y posterior a cada imagen se le hizo un recorte para solo tener el área de interés y cada imagen tuvo

una resolución promedio de 2101 x 2630 píxeles y de esa manera reducir el tiempo del algoritmo de segmentación.

3.7. Proceso de segmentación

La segmentación es un paso clave en el procesamiento de imágenes y el proceso de segmentación consiste en dividir la imagen en diferentes regiones y extraer la región objetivo de interés [28], en este trabajo se implementó el algoritmo de segmentación crecimiento de región para imágenes de color, adecuado por su forma de funcionar en el que se puede obtener una región como el fondo de las imágenes para separar los granos.

3.7.1. Crecimiento de región

Ésta técnica de segmentación es un procedimiento para agrupar píxeles en una región basado en un criterio predefinido, éste enfoque básico inicia con un pixel o conjunto de píxeles semilla y éstas regiones crecen por adición a cada pixel vecino que comparten propiedades similares a la semilla [14], la idea básica del algoritmo de crecimiento de región es agrupar píxeles con propiedades similares que conforman una región en una imagen, se requiere de un pixel semilla de referencia para que el algoritmo busque píxeles similares y forme la región. Esta técnica resulta de utilidad para segmentar el fondo y tener como resultado una matriz binaria con valores que distingue a los granos del fondo y que al procesar la información se puede utilizar la ubicación de los granos de frijol.

Para la evaluación en la búsqueda de píxeles similares al pixel semilla se utilizó distancia euclidiana ponderada para imágenes a color en el espacio de color CIE L*a*b*.

Fórmula para la medición de similitud de píxeles para el algoritmo de crecimiento de regiones:

$$D_{lab} = \sqrt{(w1 * l)^2 + (w2 * a)^2 + (w3 * b)^2}$$

El pseudocódigo 1 muestra el método de implementación del algoritmo de crecimiento de región por color.

Algoritmo 1 Algoritmo de crecimiento de región por color

```

1: procedure CRECIMIENTODEREGION(Imagen, x, y, maxdist)
2:   I ← RGB Imagen
3:   x ← posición x del pixel
4:   y ← posición y del pixel
5:   maxdist ← distancia de pixel vecino
6:   Ilab ← convertir rgb a cie lab(I)
7:   SEMILLA ← Ilab(x, y, :)
8:   w1 ← valor peso para l*
9:   w2 ← valor peso para a*
10:  w3 ← valor peso para b*
11:  [fh, fw] ← size(Ilab)
12:  svecinos ← [-10; 10; 0 - 1; 01; -11; 11; -1 - 1; 1 - 1]
13:  REGION ← (fh, fw)
14:  REGION(x, y) ← 1

```

```

15:  vecinos_pendiente ← vecinos_pendientes + [x, y]
16:  while existe vecinos pendientes do
17:    x ← vecinos pendientes(1, 1)
18:    y ← vecinos pendientes(1, 2)
19:    for svecinos = 1 to longitud(svecinos) do
20:      xn ← x + neighbours(j, 1)
21:      yn ← y + neighbours(j, 2)
22:      limite ← (xn ≤ fhandyn ≤ fw)
23:      if (limite and ( REGION(xn,yn) == 0)) then
24:        distancia ← calcularDlab
25:        if distancia < maxdist then
26:          REGION(xn, yn) ← 1
27:          vecinos pendientes ← vecinos pendientes + [xn, yn]
28:        else
29:          REGION(xn, yn) ← 0
30:        end if
31:      end if
32:    end for
33:  end while
34:  Return REGION
35: end procedure

```

La máscara de segmentación es una matriz del tamaño de la imagen a la que segmenta y está conformada de valores de 0 y 1, y las posiciones de los granos tendrán valores de 0, mientras que en las posiciones del fondo tendrán valores de 1. La Figura 3.9 muestra el resultado de aplicar segmentación por crecimiento de región, donde las manchas negras son el resultado de la segmentación y representan los granos de frijol. El pixel semilla que debe seleccionar, tiene que ser un pixel que corresponda al fondo de la imagen, el algoritmo no considera la acumulación de regiones por el hecho de aplicar la segmentación solo al fondo y no hacerlo de grano por grano. La plantilla resultante o mascarará servirá para obtener la gama de colores de los granos y serán empleados en la generación de los histogramas de color.

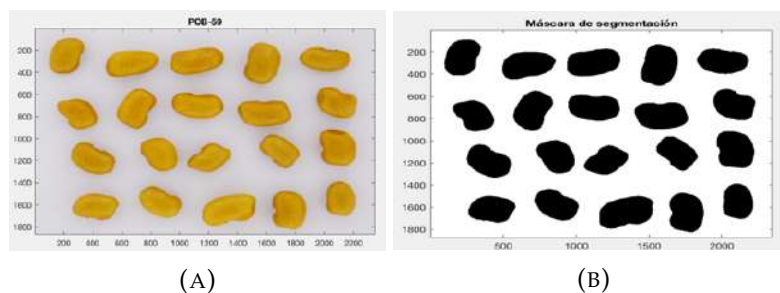


FIGURA 3.9: Ejemplo de segmentación de la población 59; A) Población 59, B) Máscara de segmentación

Es necesario hacer un almacenamiento de las máscaras de segmentación de las poblaciones de frijol, para tener una base de datos de máscaras y evitar estar aplicando el algoritmo de segmentación cada que se necesite una máscara de cierta población de granos de frijol. En la sección de granos la máscara será de gran utilidad para dividir las poblaciones en datos de

entrenamiento y prueba para la extracción de características que se utilizaran en el algoritmo de clasificación. Puede guardar cada máscara como un archivo dentro de una carpeta que las identifique como a continuación el ejemplo siguiente.

```
Máscaras
├── Pob_01
├── Pob_02
├── ⋮
```

Tendrá un conjunto de máscaras por el total de las poblaciones, con esto podrá obtener las características necesarias para la clasificación.

3.7.2. Obtención de los pesos para el algoritmo de segmentación

El valor de los pesos para cada canal del espacio de color CIE L*a*b* que se requieren para el algoritmo de segmentación se obtienen de la siguiente forma:

Para cada una de las tres imágenes de fondo que se utilizaron para contrastar con los colores de los granos. Se realizó conversión al espacio de color RGB a CIE L*a*b* y por cada componente se aplicó la siguiente fórmula para la obtención de los pesos:

$$w_i = \frac{1}{\sigma_i},$$

Donde:

w_i : es el resultado de la normalización de cada componente de color
 σ_i : es la desviación estandar del componente de color

Quedando de la forma que a continuación se muestra:

```
w1 = 1 / desviación estándar (Componente L*)
w2 = 1 / desviación estándar (Componente a*)
w3 = 1 / desviación estándar (Componente b*)
```

Con este método se obtiene el peso para cada componente de color, mismo que será utilizado para ponderar la medición de la distancia euclidiana de cada pixel que se compare con el pixel semilla, para la búsqueda de pixeles similares que formen la región de segmentación.

3.7.3. Selección de granos para entrenamiento y prueba

Para la obtención de las características y generar los histogramas se necesita de la información que ofrece el color de los granos. El proceso de selección de granos de cada población consiste en utilizar la máscara de segmentación para realizar el procedimiento de separación aleatorio de granos, y que servirá para la obtención de los valores de los pixeles, información útil para la creación de los histogramas tanto de entrenamiento como de prueba para el entrenamiento del algoritmo de clasificación.

De la máscara de segmentación resultante del algoritmo de crecimiento por región, se aplica método de etiquetado de componentes conexas que da número a cada grupo de pixeles y se realiza

una eliminación de aquellos grupos pequeños que no son de interés pero que están como resultado de la segmentación de algún elemento no apropiado como residuos de un material o polvo, a la máscara resultante con filtrado se hace un nuevo etiquetado de componentes conexas para realizar una permutación aleatoria para dividir los granos, obtenemos de esta forma dos máscaras una para obtener la información de los granos de la población para la generación de histogramas de entrenamiento y la otra contendrá los granos de prueba para la generación de histogramas de validación. El algoritmo 2 muestra la implementación para la separación aleatoria de granos de las poblaciones para entrenamiento y prueba.

Algoritmo 2 Generación de máscaras de segmentación para granos de entrenamiento y prueba

```

1: procedure MASKENTRENAMIENTOPRUEBA(Máscara de segmentación)
2:   Mask ← /* Máscara de segmentación asignada a mask */
3:   Mask ← ~ Mask /* Invertir valores de pixeles */ /* Limpieza de pixeles que no son parte
   del objeto de interés */
4:   ML ← ComponentesConectados(Mask) /* Etiquetar granos de frijol */
5:   propiedad ← propiedades(ML) /* Calcular propiedades de los objetos de la imagen */
6:   s ← busca([propiedad.Area] <1000)
7:   for i ← 1 to longitud(s) do
8:     index ← buscar(ML == s(i))
9:     Mask(index) ← 0 /* rellenar con ceros pixeles pequeños */
10:  end for /* Máscaras de segmentación para granos de frijol para entrenamiento y prueba */
11:  Mask_e ← uint8(Mask) /* Máscara de segmentación para granos de entrenamiento */
12:  Mask_v ← uint8(Mask) /* Máscara de segmentación para granos de validación */
13:  [ML, GT] ← ComponentesConectados(Mask) /* Etiquetar granos de frijol */
14:  exemplar ← 1:1:GT /* vector de 1 en 1 hasta total de granos */
15:  shufled ← exemplar(randperm(length(exemplar))) /* permutación aleatoria */
16:  N ← round(GT/2) /* División para entrenamiento y validación */
17:  trainingbeans ← shufled(1:N)
18:  for i ← 1 to longitud(trainingbeans) do
19:    index ← buscar(ML == trainingbeans(i))
20:    Mask_v(index) ← 0
21:  end for
22:  testbeans ← shufled(N+1:GT)
23:  for i ← 1 to longitud(testbeans) do
24:    index ← buscar(ML == testbeans(i))
25:    Mask_e(index) ← 0
26:  end for
27:  Mask_e ← _e
28:  Mask_v ← _v
29:  Return Mask_e, Mask_v
30: end procedure

```

A continuación, la Figura 3.10 muestra un ejemplo de la selección de granos aleatorios, se observa que al utilizar la máscara de segmentación para división de la población y tener granos de entrenamiento y granos para prueba.

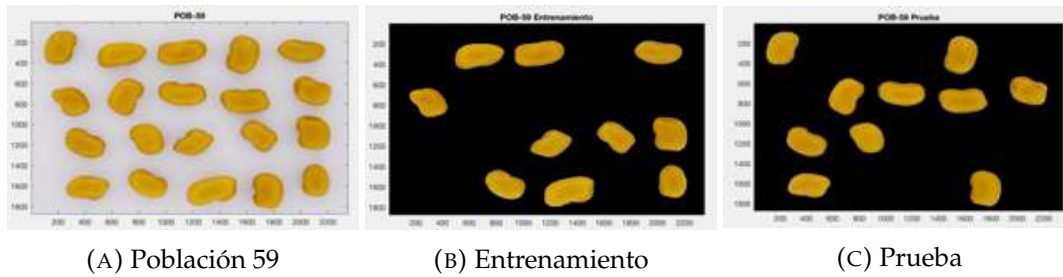


FIGURA 3.10: Ejemplo de selección de granos para entrenamiento y prueba

3.8. Histogramas

Una imagen digital esta definida como una función de 2 dimensiones $f(x, y)$ donde x y y son coordenadas espaciales de un plano, y la amplitud de f a cualquier par de coordenadas (x, y) se llama intensidad o nivel de gris de una imagen. Si x , y y la amplitud de f son valores finitos y discretos, podemos decir que es una imagen digital y que esta compuesta por un número finito de elementos llamados pixeles y cada uno tiene una ubicación y un valor particular. La representación de una imagen es mostrada en la Figura 3.11, que se representa como una matriz de $N \times M$ teniendo el origen $(0, 0)$ y el siguiente valor se representan como $(x, y) = (0, 1)$ y corresponde a la segunda muestra en la primera fila.

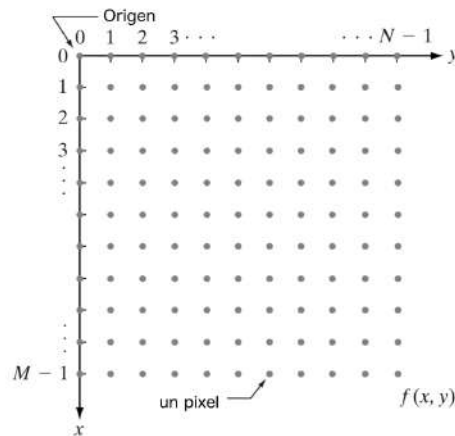


FIGURA 3.11: Representación digital de una imagen

Al representar como una matriz de $N \times M$ y tener coordenadas $f(x, y)$ y se puede tener una notación matricial como lo muestra la Figura 3.12, $a_{ij} = f(x = i, y = j) = f(i, j)$ que son matrices idénticas.

Sea Z y R el conjunto de los enteros reales y el conjunto de los números reales respectivamente. El proceso de muestreo puede ser visto en el plano x, y dentro de una cuadrícula, con las coordenadas de cada cuadrícula como un par de elementos del producto cartesiano Z^2 , que es el conjunto de todos los pares ordenados de elementos (z_i, z_j) , con z_i y z_j son números enteros de Z . Por lo tanto, $f(x, y)$ es una imagen digital si (x, y) son números enteros de Z^2 y f es una función que asigna un valor de nivel de gris (es decir, un número real del conjunto de números

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0, N-1) \\ f(1,0) & f(1,1) & \cdots & f(1, N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1, N-1) \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,N-1} \\ a_{1,0} & a_{1,1} & \cdots & a_{1,N-1} \\ \vdots & \vdots & & \vdots \\ a_{M-1,0} & a_{M-1,1} & \cdots & a_{M-1,N-1} \end{bmatrix}$$

FIGURA 3.12: Notación matriz compacta y tradicional

reales R) a cada par de coordenadas distintas (x, y) . Si los niveles de grises también son enteros, Z reemplaza a R , y una imagen digital se convierte entonces en una función de 2 dimensiones cuyas coordenadas y valores de amplitud son enteros.

$$L = 2^k$$

El número de niveles de gris comúnmente es un entero elevado a la potencia 2, una imagen puede tener k niveles de grises [14].

En imágenes a escala de gris, cada pixel está representado por un valor de intensidad o nivel de gris que va de [0-255], el 0 significa que no hay luz y representa el color negro, el valor máximo en el rango que comúnmente se usa es 255 significa que se aplica la cantidad de luz máxima posible produciendo color blanco. Una imagen de color es un conjunto de pixeles de diferentes tonalidades que se mezclan, el color de cada pixel se forma con varios valores tomados de los colores primarios mezclados para lograr el color.

El histograma de una imagen se puede representar como la frecuencia de diferentes niveles de gris. Cada valor en el histograma de una imagen representa un nivel de gris. 0 que representa el negro se muestra a la izquierda y 255 que representa el blanco se muestra a la derecha.

3.8.1. Histograma 1D

El histograma de una imagen digital con niveles de gris está en el rango $[0, L - 1]$ como una función discreta (3.7)

$$h(r_k) = n_k \quad (3.7)$$

Donde:

r_k : es el k_{th} nivel de gris,

n_k : es el número de pixeles en la imagen teniendo el nivel de gris r_k

La normalización de un histograma esta dada por (3.8):

$$p(r_k) = \frac{n_k}{n} \quad (3.8)$$

Donde:

n : es el número de píxeles en la imagen

La Figura 3.13 contiene un ejemplo de histograma, en el eje X se tiene a los niveles de gris y en el eje Y se tiene las frecuencias de cada pixel.

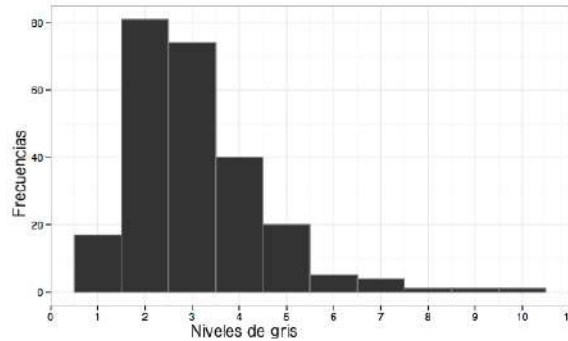


FIGURA 3.13: Histograma

3.8.2. Distribución de probabilidad conjunta

Si X y Y son dos variables aleatorias discretas, la distribución de probabilidad para sus ocurrencias simultáneas pueden ser representadas por una función con valores $f(x, y)$ para cualquier par de valores (x, y) dentro del rango de las variables aleatorias X y Y .

La **distribución de probabilidad conjunta** o **función de masa de la probabilidad** de dos variables aleatorias discretas X y Y está definida como:

$$f_{XY}(x, y) = P(X = x, Y = y)$$

que es el valor de $f_{XY}(x, y)$ de la probabilidad de ocurrencia de x y y al mismo tiempo, donde la coma significa "y"

$$f_{XY}(x, y) = P(X = x, Y = y) = P(X = x) y (Y = y)$$

Se puede definir el rango conjunto para X y Y como:

$$R_{XY} = \{(x, y) | P_{XY}(x, y) > 0\}$$

En particular, si $R_X = \{x_1, x_2, \dots\}$ y $R_Y = \{y_1, y_2, \dots\}$, entonces podemos escribir

$$R_{XY} \subset R_X \times R_Y = \{(x, y) | x_i \in R_X, y_i \in R_Y\}$$

Se define $R_{XY} = R_X \times R_Y$ para simplificar el análisis, además algunos pares (x_i, y_j) en $R_X \times R_Y$, $P_{XY}(x_i, y_j)$ puede ser cero. Para dos variables discretas aleatorias X y Y se tiene:

$$\sum_{(x_i, y_i) \in R_{XY}} P_{XY}(x_i, y_i) = 1$$

Se puede usar la función de masa de la probabilidad conjunta para buscar $P((X, Y) \in A)$ para cualquier conjunto $A \subset \mathbb{R}^2$. Específicamente tenemos:

$$P((X, Y) \in A) = \sum_{(x_i, y_i) \in (A \cap R_{XY})} P_{XY}(x_i, y_i)$$

El evento $X = x$ puede ser escrito como $\{(x_i, y_i) : x_i = x, y_j \in R_Y\}$. También, el evento $Y = y$ puede ser escrito como $\{(x_i, y_i) : x_i \in R_X, y_j = y\}$. Así, podemos escribir

$$P_{XY}(x, y) = P(X = x, Y = y) = P((X = x) \cap (Y = y))$$

En base a la definición anterior [2] podemos generar histogramas en dos dimensiones formados por la probabilidad conjunta de variables aleatorias discretas, la imagen 3.14 muestra un ejemplo de distribución de probabilidad conjunta expresada en histograma, donde las frecuencias de un valor aleatorio discreto con respecto a otro dan forma de campana gaussiana representada al centro en la imagen del ejemplo.

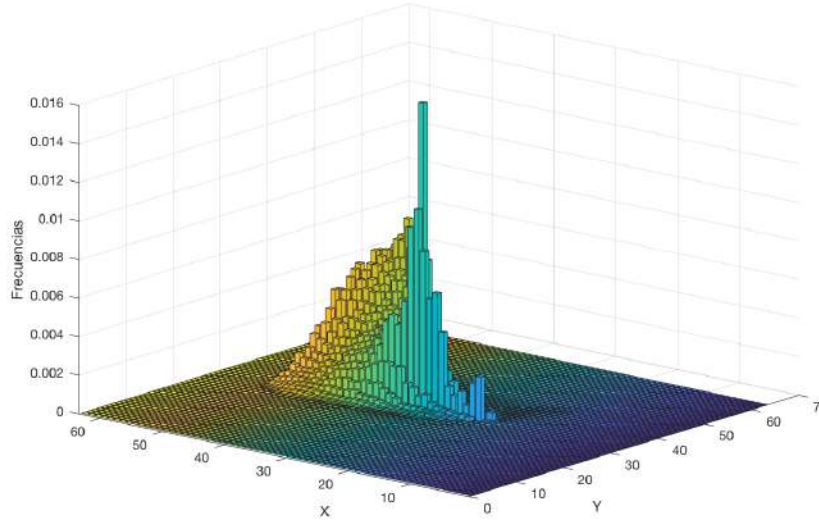


FIGURA 3.14: Histograma de probabilidad conjunta

3.8.3. Histograma de 2 dimensiones

En base a la distribución de probabilidad conjunta de los píxeles de cada canal del espacio de color de una imagen se puede generar histograma en 2 dimensiones como una matriz de $L \times L$, y la generación de histogramas en 2 dimensiones para una imagen ocupando los canales de cromaticidad para HSI y CIE $L^*a^*b^*$ esta dada por la probabilidad conjunta de los canales HS y la probabilidad conjunta de los canales a^*b^* que se describe como:

El histograma en 2 dimensiones es una matriz de $L \times L$, representada por pares de $[(x, y), (x, y)]$, para una imagen sea r_{ij} el número total de ocurrencias de un par (i, j) el cual representa el píxel (x, y) con $f(x, y) = i$ y $g(x, y) = j$.

La probabilidad conjunta de p_{ij} está dada por (3.9):

$$p_{ij} = \frac{r_{ij}}{N^2} \quad (3.9)$$

Donde:

$$i, j = [1 - L]$$

$$\sum_{i=1}^L \sum_{j=1}^L p_{ij} = 1$$

p_{ij} es el histograma en 2 dimensiones de una imagen, sea N^2 la cantidad de $r_{i,j}$ [13].

Ejemplo de histograma en 2 dimensiones lo muestra la Figura 3.15, donde se tiene la frecuencia de la probabilidad conjunta de los píxeles de un canal con respecto a otro canal. Al observar la imagen se nota la diferencia de histogramas de cada espacio de color, los histogramas del espacio

de color HSI son frecuencias que se dispersan en el espacio de frecuencias del histograma, a diferencia de ese espacio de color, los histogramas del espacio de color CIE L*a*b* las frecuencias se agrupan más al centro dando distintas amplitudes y son más descriptivas, basado en el hecho que se nota la frecuencia en formas de campanas que corresponde a cada conjunto de tonos de color. Por ejemplo, si tenemos tres distintos colores de granos e impera un color este se notará con más frecuencia en el histograma, siendo más alta que las demás frecuencias y se tendrá tres cúspides con distintos tamaños y longitudes.

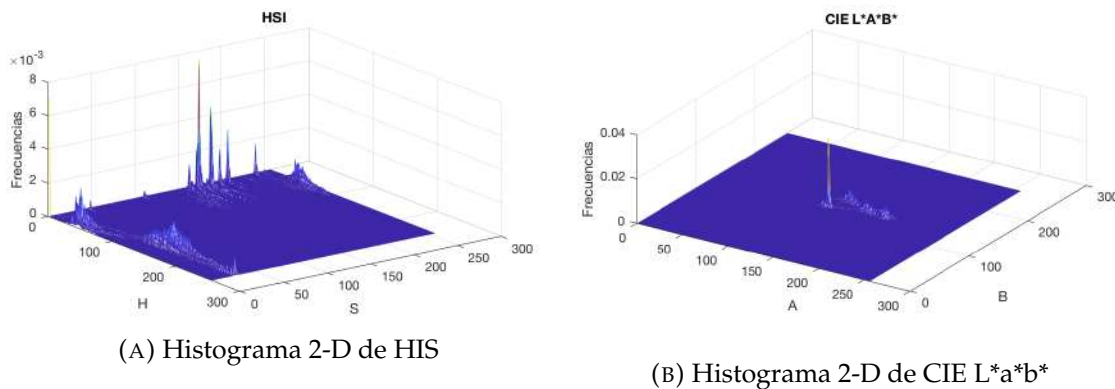


FIGURA 3.15: Ejemplo de histograma en 2-D

3.8.4. Histogramas en 3 dimensiones

El mismo principio se aplica para la generación de histogramas en 3 dimensiones basado en la distribución de probabilidad conjunta, se considera una tercera variable discreta para tener la ocurrencia de los elementos.

Se puede generalizar a más variables aleatorias discretas como la siguiente definición:

Si X_1, X_2, \dots, X_n son variables aleatorias discretas entonces su función de masa de probabilidad conjunta es:

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

La densidad conjunta para n variables aleatorias es una función en \mathbb{R}^n . no es cero solo un conjunto finito y contable de puntos en \mathbb{R}^n . si sumamos estos puntos obtenemos 1.

$$\sum_{x_1, \dots, x_n} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = 1$$

3.8.5. Extracción de histogramas

A partir de las características obtenidas de los granos de frijol de cada población ocupando los espacios de color HSI y CIE $L^*a^*b^*$ y con base a la probabilidad conjunta se generaron histogramas en 2 dimensiones e histogramas en 3 dimensiones de las 54 poblaciones de frijol, con la finalidad de saber si es necesaria la carga computacional que involucra integrar el tercer componente que aporta la intensidad a la cromaticidad y en base a eso tener más de un elemento que permite saber los porcentajes de precisión de clasificación para la identificación de acuerdo a la variedad local y coloración del grano.

La forma de extracción de las características colorimétricas de cada grano fue mediante la máscara de segmentación de entrenamiento y de prueba donde las posiciones con valores 0 de la matriz de la máscara corresponden a las posiciones de los píxeles de los granos de la población, en el siguiente algoritmo 3 muestra el procedimiento empleado para la extracción de los valores de los píxeles y creación de los histogramas en 2 dimensiones del espacio de color HSI.

En el algoritmo los canales correspondientes a la cromaticidad que son matiz y saturación de HSI son empleados para la generación de histogramas en 2 dimensiones y en caso del espacio de color HSI la fórmula de conversión da resultados normalizados en [0-1]. Para trabajar a una escala de 256 bins es necesario el escalado de [0-1] a [1-256]. El algoritmo 3 desglosa el procedimiento empleado en la generación del histograma.

Algoritmo 3 Generación de histogramas en 2-D de HSI a 256 bins

```

1: procedure HIST2DHS(imagen,Máscara de segmentación)
2:   I ← imagen
3:   Mask ← Máscara de segmentación
4:   [global_rows, global_cols] ← Dimensiones(I) /* Obtenemos las dimensiones de la
   imagen */
5:   MaskV ← reestructura(Mask, global_rows * global_cols, 1) /* Reestructurar matriz de
   máscara para obtener posiciones de píxeles */
6:   indices ← busca(MaskV == 0) /* Devuelve todos los índices de los píxeles de los granos
   */
7:   img ← Convertir rgb a hsi(I)
8:   data_raw ← reestructura(img, global_rows * global_cols, 3) /* Reestructurar para
   trabajar con los canales del espacio de color */
9:   datosHSI ← data_raw(indices,:)
10:  n ← Dimensiones(datosHSI) /* Devuelve el total de píxeles expresado en n */ /* Escalado
   y desplazamiento, rango de 1/256, depende del lenguaje de programación */
11:  datosHSI ← (datosHSI * 255) + 1 Escalado a [0-255] más desplazamiento [1-256]
12:  datosHSI ← redondear(datosHSI)
13:  indexH ← unicos(datosHSI(:,1)) /* Canal H */
14:  indexS ← unicos(datosHSI(:,2)) /* Canal S */
15:  h2dhs ← ceros(256,256)
16:  for i = 1 to longitud(indexH) do
17:    for j = 1 to longitud(indexS) do /* Conjunta de los 2 canales */
18:      ind ← busca(indexH(i) == datosHSI(:,1) and
19:        indexS(j) == datosHSI(:,2))
20:      h2dhs(indexH(i),indexS(j)) ← longitud(ind) /* Asignación de las frecuencias */

```



```

21:     end for
22: end for
23: h2dhs ← h2dhs / n /* Normalización */
24: Return h2dhs
25: end procedure

```

El procedimiento para la generación de histogramas en 2 dimensiones para el espacio de color CIE L*a*b* es muy similar solo cambia la forma que debe darse al tratamiento adecuado para el escalado, la conversión de RGB a CIE L*a*b* deja los canales en sus valores de rango L* de [0-100], a* y b* de [-128-129], se debe realizar normalización en L* y desplazamiento en a* y b*. observe el algoritmo 4.

Algoritmo 4 Generación de histograma 2-D de a* y b* de CIE L*a*b*

```

1: procedure HIST3DLAB(imagen,Máscara de segmentación)
2:   I ← Imagen
3:   Mask ← Máscara de segmentación
4:   [global_rows, global_cols] ← Dimensiones(I) /* Obtenemos las dimensiones de la
   imagen */
5:   MaskV ← reestructura(Mask, global_rows * global_cols, 1) /* Reestructurar matriz de
   máscara para obtener posiciones de pixeles */
6:   indices ← busca(MaskV == 0) /* Devuelve todos los indices de los pixeles de los granos
   */
7:   img ← Convertir rgb a lab(I)
8:   data_raw ← reestructura(img, global_rows * global_cols, 3) /* Reestructurar para
   trabajar con los canales del espacio de color */
9:   datosHSI ← data_raw(indices,:)
10:  n ← Dimensiones(datosHSI) /* Devuelve el total de pixeles expresado en n */
11:  datosLAB(:,1) ← datosLAB(:,1) / 100 /* normalización [0-1] */
12:  datosLAB(:,1) ← datosLAB(:,1) * 255 /* escalado [0-255] */
13:  datosLAB(:,1) ← datosLAB(:,1) + 1 /* desplazamiento [1-256] */
14:  datosLAB(:,2:3) ← datosLAB(:,2:3) + 129 /* Desplazamiento de [-128-127] a [1-256] a* b* */
15:  datosLAB ← redondear(datosLAB)
16:  indexA ← unicos(datosLAB(:,2)) /* canal a* */
17:  indexB ← unicos(datosLAB(:,3)) /* canal b* */
18:  h2dlab ← ceros(256,256)
19:  for i = 1 to longitud(indexA) do
20:    for j = 1 to longitud(indexB) do /* Conjunta de los 2 canales */
21:      ind ← busca(indexA(i) == datos(:,1) and
22:                indexB(j) == datos(:,2))
23:      h2dlab(indexA(i),indexB(j)) ← longitud(ind) /* Asignación de las frecuencias */
24:    end for
25:  end for
26:  h2dlab ← h2dlab / n /* Normalización */
27:  Return h2dlab
28: end procedure

```

Trabajar la generación de histogramas en 3 dimensiones a [0-255] es un proceso computacionalmente caro y para este trabajo se redujo el rango de pixeles a [0-127], crear

histogramas a 128 niveles de intensidad con probabilidad conjunta de los píxeles de tres componentes de color se mantiene la misma información a una escala distinta, lo que se traduce en mantener la información en forma compacta al no tenerla a un rango de [0-255], esto reduce el tiempo de generación tanto para los espacios de color HSI como CIE L*a*b*, en total se generaron 54 histogramas para entrenamiento y 54 histogramas de prueba de las 54 poblaciones nativas de frijol ocupadas en este trabajo.

El algoritmo 5 siguiente muestra el procedimiento de generación de histogramas y se utilizó la probabilidad conjunta de los tres componentes de color de cada espacio de color propuesto como descriptor.

Algoritmo 5 Generación de histogramas en 3-D de HSI a 128 bins

```

1: procedure HIST3DHSI(imagen,Máscara de segmentación)
2:   I ← Imagen
3:   Mask ← Máscara de segmentación
4:   [global_rows,global_cols] ← Dimensiones(I) /* Obtenemos las dimensiones de la
   imagen */
5:   MaskV ← reestructura(Mask, global_rows * global_cols, 1) /* Reestructurar matriz de
   máscara para obtener posiciones de píxeles */
6:   indices ← busca(MaskV == 0) /* Devuelve todos los índices de los píxeles de los granos
   */
7:   img ← Convertir rgb a hsi(I)
8:   data_raw ← reestructura(img, global_rows * global_cols, 3) /* Reestructurar para
   trabajar con los canales del espacio de color */
9:   datosHSI ← data_raw(indices,:)
10:  n ← Dimensiones(datosHSI) /* Devuelve el total de píxeles expresado en n */ /* Escalado
   y desplazamiento, rango de 1/128, depende del lenguaje de programación */
11:  datosHSI ← (datosHSI * 127) /* [0-1] escalado a [0-127] */
12:  datosHSI ← datosHSI + 1 /* desplazamiento de 1 para [1-128] */
13:  datosHSI ← redondear(datosHSI)
14:  indexH ← unicos(datosHSI(:,1)) /* Canal H */
15:  indexS ← unicos(datosHSI(:,2)) /* Canal S */
16:  indexI ← unicos(datosHSI(:,3)) /* Canal I */
17:  h3hsi ← ceros(128,128,128)
18:  for i = 1 to longitud(indexH) do
19:    for j = 1 to longitud(indexS) do
20:      for k = 1 to longitud(indexI) do /* Conjunta de los 3 canales */
21:        ind ← busca(indexH(i) == datosHSI(:,1) and
22:                    indexS(j) == datosHSI(:,2) and
23:                    indexI(k) == datosHSI(:,3))
24:        h3hsi(indexH(i),indexS(j),indexI(k)) ← longitud(ind) /* Asignación de las
   frecuencias */
25:      end for
26:    end for
27:  end for
28:  h3hsi ← h3hsi / n /* Normalización */
29:  Return h3hsi

```

30: **end procedure**

Un procedimiento similar para la generación de histogramas en 3 dimensiones para el espacio de color CIE L*a*b* como lo muestra el algoritmo 6.

Algoritmo 6 Generación de histograma 3-D de CIE L*a*b* a 128 bins

```

1: procedure HIST3DLAB(imagen,Máscara de segmentación)
2:   I ← Imagen
3:   Masks Máscara de segmentación
4:   [global_rows, global_cols] ← Dimensiones(I) /* Obtenemos las dimensiones de la
   imagen */
5:   MaskV ← reestructura(Mask, global_rows * global_cols, 1) /* Reestructurar matriz de
   máscara para obtener posiciones de pixeles */
6:   indices ← busca(MaskV == 0) /* Devuelve todos los indices de los pixeles de los granos
   */
7:   img ← Convertir rgb a lab(I)
8:   data_raw ← reestructura(img, global_rows * global_cols, 3) /* Reestructurar para
   trabajar con los canales del espacio de color */
9:   datosHSI ← data_raw(indices,:)
10:  n ← Dimensiones(datosHSI) /* Devuelve el total de pixeles expresado en n */
11:  datos(:,1) ← datos(:,1) / 100 /* normalización [0-1] */
12:  datos(:,1) ← datos(:,1) * 127 /* escalado [0-127] */
13:  datos(:,1) ← datos(:,1) + 1 /* desplazamiento [1-128] */
14:  datos(:,2:3) ← datos(:,2:3) + 128 /* desplaza de [-128 - 127] a [0-255] a* b* */
15:  datos(:,2:3) ← datos(:,2:3) / 255 /* normalización [0-1] de a* b* */
16:  datos(:,2:3) ← datos(:,2:3) * 127 /* escalado [0-127] */
17:  datos(:,2:3) ← datos(:,2:3) + 1 /* desplazamiento de 1 [1-128] */
18:  datos ← redondear(datos)
19:  indexL ← unicos(datos(:,1)) /* canal L* */
20:  indexA ← unicos(datos(:,2)) /* canal a* */
21:  indexB ← unicos(datos(:,3)) /* canal b* */
22:  h3lab ← ceros(128,128,128)
23:  for i = 1 to longitud(indexL) do
24:    for j = 1 to longitud(indexA) do
25:      for k = 1 to longitud(indexB) do /* Conjunta de los 3 canales */
26:        ind ← busca(indexL(i) == datos(:,1) and
27:                    indexA(j) == datos(:,2) and
28:                    indexB(k) == datos(:,3))
29:        h3lab(indexL(i),indexA(j),indexB(k)) ← longitud(ind) /* Asignación de las
   frecuencias */
30:      end for
31:    end for
32:  end for
33:  h3lab ← h3lab / n /* Normalización */
34:  Return h3lab
35: end procedure

```

3.9. Paleta de colores

Como parte de la necesidad de identificación de poblaciones de acuerdo a la variedad local se requiere clasificación de poblaciones nativas de frijol por color del grano para tener información que ayude a saber que poblaciones reúnen características colorimétricas similares. Con lo que se requiere tener una referencia de color y para lograr la clasificación de poblaciones nativas de frijol por coloración se elaboró una paleta de colores personalizada que toma como muestra de color los granos de las poblaciones para formar grupos de color y dar una etiqueta a cada población que la describa por color.

Para la elaboración fue necesario utilizar las imágenes que se obtuvieron en el proceso de adquisición de imágenes. Se ocupó la máscara de segmentación para buscar las áreas correspondientes a los granos, por cada grano de frijol fue generada una imagen con el nombre de la población a la que pertenece más un número adicional relativo al número consecutivo dentro de su población, fueron agrupadas las imágenes de los granos de similar coloración, se realizó una ordenación de mayor a menor saturación y como resultado se obtuviera la categoría de color conformada por granos de frijol como imagen.

Se obtiene un listado de los nombres de los granos agrupados para obtener el nombre de la población a la que pertenecen y de esa forma conformar el nombre que describe por color a la población, si una población tiene granos en distintos grupos su nombre será una combinación de los nombres de cada grupo de color, como lo muestra el cuadro 3.2. si una población contiene granos de dos colores como por ejemplo la población 03 tendrá una etiqueta descriptiva de “Mezcla negro y rojo” consulte apéndice B para ver imagen de la población 03.

El propósito de tener una referencia colorimétrica fue para dar descripción por color a cada una de las 54 poblaciones de frijol, ello permite tener una etiqueta descriptiva colorimétrica útil para la identificación de poblaciones nativas de frijol por similitud de color. Con todos los granos de las 54 poblaciones se logró elaborar 12 muestras de color Figura 3.16.

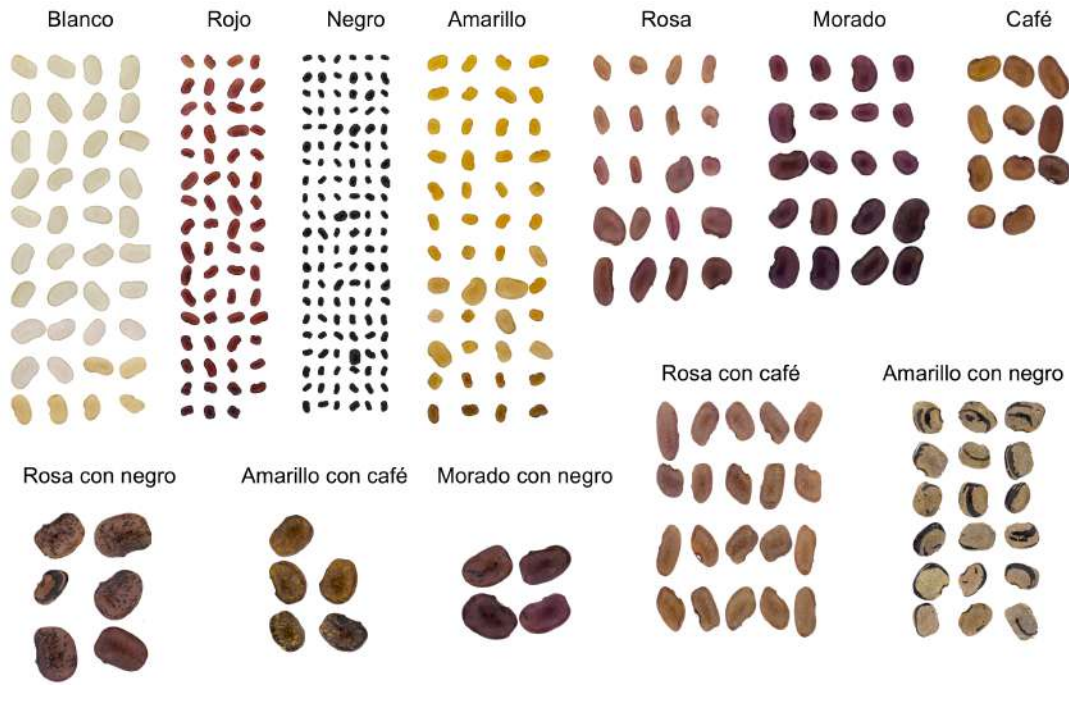


FIGURA 3.16: Paleta de colores

Basada en la paleta de colores se elaboró la descripción a cada población y se obtuvo 21 distintas clases para el conjunto de 54 poblaciones, a continuación, en el Cuadro 3.2 se listan las distintas clases por descripción de color. Consulte el apéndice C si desea ver la descripción por color dada a cada población de frijol empleada en este trabajo.

CUADRO 3.2: Taxonomía de color para poblaciones nativas de frijol.

Núm.	Taxonomía de color	Ocurrencias
1	Amarillo	4
2	Amarillo con negro	1
3	Blanco	4
4	Mezcla amarillo con negro, café, negro, rojo y rosa con negro	1
5	Mezcla amarillo y amarillo con negro	1
6	Mezcla amarillo y rojo	1
7	Mezcla amarillo, amarillo con café y negro	1
8	Mezcla amarillo, morado, morado con negro, negro y rosa	1
9	Mezcla amarillo, negro y rojo	2
10	Mezcla amarillo, negro y rosa	2
11	Mezcla blanco y negro	1
12	Mezcla blanco y rojo	1
13	Mezcla blanco, negro y rojo	1
14	Mezcla café y rojo	1
15	Mezcla morado y negro	3

Continúa en la siguiente página →

Cuadro 3.2 – \leftrightarrow Complementa la página anterior

Núm.	Taxonomía de color	Ocurrencias
16	Mezcla morado, morado con negro, rosa y rosa con negro	1
17	Mezcla negro y rojo	3
18	Mezcla negro, rojo y rosa	1
19	Negro	15
20	Rojo	8
21	Rosa con café	1
Número total de poblaciones nativas de frijol		54

Las 21 categorías distintas obtenidas de las 54 poblaciones reflejan la gran variabilidad de colores que las conforman.

3.10. Aprendizaje automático

Para resolver un problema en una computadora se requiere de un algoritmo que puede procesar algo mediante una entrada y una salida, por ejemplo se puede crear un algoritmo de ordenamiento que tenga como entrada un conjunto de números y como salida una lista ordenada de los números. Para algunas tareas no se tiene un algoritmo y con los avances de la tecnología la mayoría de los dispositivos de adquisición de datos son ahora digitales y registran datos fiables, además, la capacidad de almacenamiento ha crecido y el procesamiento de grandes cantidades de datos almacenados pueden ser útiles en el momento en el que se analizan y se convierten en información útil para hacer predicciones, crear un proceso que explique los datos que se observan y construir una aproximación útil que explique parte de los datos y es función del proceso de aprendizaje automático. Los patrones pueden ayudar a entender un proceso y la clasificación puede usarse para hacer predicciones, el aprendizaje automático ayudó a encontrar soluciones a muchos problemas en visión, reconocimiento de voz y robótica [4].

3.10.1. K vecinos más cercanos

El algoritmo K vecinos más cercanos es uno de los más sencillos de aprendizaje automático. Basado en su simplicidad es una técnica ampliamente utilizada y aplicada en un gran número de dominios. La versión básica de K vecino más cercano asume que todas las instancias corresponden a puntos n -dimensionales del espacio \mathbb{R}^n . La instancia de un vecino cercano es definida en términos de la distancia euclidiana. Más precisamente que una instancia arbitraria sea descrita por el vector $\langle x, f(x) \rangle$, si f es verdadera la función concepto que da, corresponde a la clase del valor $f(x)$ para cada instancia x , La instancia x puede ser descrita como el vector característico:

$$\langle a_1(x), a_2(x), \dots, a_n(x) \rangle$$

donde $a_r(x)$ denota el valor de r th atributo de instancia x . Por la distancia euclidiana estándar, $d(x_i, x_j)$ está definida por la ecuación:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

El aprendizaje para vecinos más cercanos la función concepto puede ser para valores discretos o valores reales, el algoritmo de K vecinos cercanos es descrito en el algoritmo 7.

Algoritmo 7 El algoritmo básico K vecinos más cercanos para aproximar a las funciones de valor discreto.

1: Algoritmo de entrenamiento:

2: - Para cada ejemplo de entrenamiento $\langle x, f(x) \rangle$, agregar el ejemplo a la lista ejemplos_entrenamiento

Algoritmo de clasificación:

3: -Dada una instancia de consulta x_q para ser clasificada,

4: Sea que x_1, \dots, x_k denote la k instancias de ejemplos_entrenamiento que está más cerca de x_q

5: Retornar

6:

$$\hat{f}(x_q) = \arg \max_{v \in V} \sum_{i=1}^k \delta(v, f(x_i)) \quad (3.10)$$

7:

8: donde $\delta(a, b) = 1$ si $a = b$ y $\delta(a, b) = 0$ en cualquier otro caso.

El valor de $\hat{f}(x_q)$ retornado por el algoritmo es una estimación de $f(x_q)$ es solo el modo de la función concepto f entre ejemplos de entrenamiento k más cercanos a x_q . Si $K = 1$, entonces el algoritmo de 1 vecino más cercano asigna a $\hat{f}(x_q)$ el valor $f(x_q)$ donde x_i es la instancia de entrenamiento vecina a x_q . Para valores largos de k , el algoritmo asigna el modo entre los ejemplos de entrenamiento k vecino cercano.

3.10.2. Algoritmo K vecinos más cercanos con distancia ponderada

Un refinamiento al algoritmo básico K vecinos más cercanos es ponderar la contribución del peso de cada uno de los k vecinos acorde a la distancia de la instancia x_q , dando mayor peso a los vecinos más cercanos. Dos enfoques ampliamente utilizados para calcular los pesos son: 1) Ponderar el voto de cada vecino acorde al cuadrado inverso de la distancia con x_q , y 2) Ponderar el voto de cada vecino acorde a su similitud con x_q , donde la similitud está definida como $1 - d(x_q, x_i)$ Esto puede ser complementado reemplazando la línea final del algoritmo 7 por:

$$\hat{f}(x_q) = \max_{v \in V} \sum_{i=1}^k \omega_i \delta(v, f(x_i)) \quad (3.11)$$

donde:

$$\omega_i = \frac{1}{d(x_q, x_i)^2} \quad (3.12)$$

ó

$$\omega_i = 1 - d(x_q, x_i) \quad (3.13)$$

en el caso de la ecuación 3.12, en orden de acomodo para el caso donde el punto de consulta x_q exactamente empata una de las instancias de entrenamiento x_i y el denominador $d(x_q, x_i)^2$ es cero, en este caso $\hat{f}(x_q)$ puede ser asignada a $f(x_q)$. Si son varios ejemplos de entrenamiento de este tipo, $\hat{f}(x_q)$ puede asignarse a la clasificación mayoritaria entre ellos [27].

3.10.3. Clasificación de histogramas con K vecinos más cercanos con distancia ponderada

El clasificador de aprendizaje automático toma como entrada las distribuciones de probabilidad de los colores expresada en histograma de 2 y 3 dimensiones. Se generó dos conjuntos, uno de 54 histogramas para entrenamiento y otro de 54 histogramas para prueba, con dos distintos tipos de etiquetas de clase, el primer tipo de etiqueta de clase consiste en el nombre que identifica a cada población teniendo un total de 54 etiquetas distintas para clasificación y la segunda ocupando como etiqueta la descripción de color que fue creada con la paleta de colores, con un total de 21 distintas etiquetas de clase, y se entrenó el algoritmo de aprendizaje automático para obtener un modelo que permita clasificar la población de frijol a partir de sus distribuciones de color.

El algoritmo de K-Vecinos más cercanos procesa tuplas de atributos, la información obtenida en este trabajo ha sido mediante histogramas que son matrices de 2 y 3 dimensiones y para la clasificación de la información en este trabajo se realizó la conversión de matrices a vectores, que serán los casos con atributos para clasificación.

Para la clasificación se utilizó como métrica de similitud cityblock para la comparación de la similitud de los histogramas y por dar mejores resultados 3.14 y con distancia ponderada cuadrado inverso para tener prioridad sobre los vecinos locales.

$$d_{st} = \sum_{j=1}^n |x_{sj} - x_{tj}| \quad (3.14)$$

Capítulo 4

Resultados

Con la finalidad de tener resultados de clasificación y con base a la anterior metodología se procedió a la generación de resultados, después de la adquisición y segmentación de las imágenes con la máscara de segmentación se realizó la separación aleatoria de los 20 granos de cada población para formar dos conjuntos de 10 granos cada uno y por cada conjunto se generaron histogramas de 2 o 3 dimensiones utilizando los espacios de color HSI y CIE $L^*a^*b^*$ con el objetivo de tener histogramas de entrenamiento y prueba para el algoritmo de aprendizaje automático, se generó una base de datos que contuviera los histogramas y etiquetas de las clases.

A fin de tener un concentrado de resultados se generó 20 bases de datos, a continuación se muestra la estructura de la base de datos con histogramas en 2 dimensiones.

- 54 histogramas de entrenamiento 2 dimensiones H y S espacio de color HSI
- 54 histogramas de entrenamiento 2 dimensiones a^* y b^* espacio de color CIE $L^*a^*b^*$
- 54 Histogramas de prueba 2 dimensiones H y S espacio de color HSI
- 54 Histogramas de prueba 2 dimensiones a^* y b^* espacio de color CIE $L^*a^*b^*$
- 54 etiquetas variedad local
- 54 etiquetas color del grano

El mismo procedimiento y estructura se ocupó para la generación de 20 bases de datos con la información de histogramas en 3 dimensiones.

- 54 histogramas de entrenamiento 3 dimensiones espacio de color HSI
- 54 histogramas de entrenamiento 3 dimensiones espacio de color CIE $L^*a^*b^*$
- 54 Histogramas de prueba 3 dimensiones espacio de color HSI
- 54 Histogramas de prueba 3 dimensiones espacio de color CIE $L^*a^*b^*$
- 54 etiquetas variedad local
- 54 etiquetas color del grano

4.1. Resultados de clasificación espectrofotómetro

Con el propósito de validar y comparar resultados de clasificación utilizando el método aquí propuesto y el método tradicional se obtuvieron 20 mediciones puntuales con el espectrofotómetro de cada población nativa de frijol y para la evaluación se tomaron aleatoriamente la mitad de mediciones para entrenamiento y la otra mitad para validación del algoritmo de aprendizaje automático k vecinos cercanos, se obtuvo resultados de clasificación para la identificación de 54 poblaciones nativas de frijol, el procedimiento se repitió 20 veces para la generación del

CUADRO 4.1: Resultados de clasificación variedad local con valores de medición de espectrofotómetro utilizado el algoritmo de clasificación K vecinos más cercanos

Ejecuciones	K=1	K=3	K=5	K=7	K=9	K=11	K=13	K=15
1	51.57	51.93	52.85	53.22	53.78	53.22	53.78	54.14
2	50.83	53.04	53.41	53.78	54.33	53.96	54.70	54.51
3	51.57	54.70	53.96	53.41	54.14	53.78	54.70	54.70
4	50.83	55.25	54.51	55.25	54.70	54.51	54.14	53.59
5	51.01	53.78	55.06	53.96	53.59	53.78	53.96	53.78
6	51.20	50.46	51.38	51.01	51.38	51.38	51.57	51.75
7	52.67	53.59	52.67	53.41	53.41	53.41	53.22	53.04
8	53.04	52.12	54.14	54.33	54.70	54.88	55.06	55.06
9	49.36	50.28	50.83	51.38	49.91	50.46	50.83	51.38
10	53.78	55.80	55.99	55.06	56.17	55.80	55.62	55.62
11	51.57	53.04	52.85	53.04	53.41	52.85	52.67	52.85
12	51.20	53.96	54.88	54.70	53.41	53.22	53.22	53.41
13	52.30	53.22	53.04	53.96	54.14	53.59	53.96	54.33
14	49.17	52.12	53.96	54.70	54.51	53.96	54.70	53.59
15	50.46	50.09	51.01	50.64	51.57	51.20	51.01	50.83
16	50.46	51.75	51.57	52.12	52.30	51.93	52.30	52.30
17	51.38	51.75	52.67	52.85	52.30	53.04	52.67	52.67
18	54.14	53.96	56.17	55.06	55.06	55.06	54.88	54.70
19	52.12	51.93	54.33	54.51	54.51	54.88	55.06	54.70
20	52.12	53.41	53.41	52.85	53.59	53.59	53.59	53.78
Media	51.54	52.81	53.43	53.46	53.55	53.43	53.58	53.54
Desviación estándar	1.27	1.58	1.52	1.35	1.45	1.36	1.38	1.28

concentrado de resultados de clasificación de poblaciones nativas de frijol por su variedad local y por color del grano.

Los resultados de clasificación por variedad del grano obtenidos con las mediciones puntuales las precisiones son similares y no varían conforme se indica un mayor número de vecinos cercanos en el parámetro K del algoritmo de clasificación K vecinos más cercanos.

Para la clasificación de poblaciones nativas de frijol por color del grano para mediciones puntuales se ocuparon las mismas etiquetas descriptivas por población que se generaron con la paleta de color y que también son ocupadas para la clasificación de las poblaciones con histogramas del método propuesto los resultados en el cuadro 4.2.

CUADRO 4.2: Resultados de clasificación por color del grano espectrofotómetro

Ejecuciones	K=1	K=3	K=5	K=7	K=9	K=11	K=13	K=15
1	70.17	72.56	73.66	73.85	74.03	74.22	74.77	73.66
2	69.06	70.17	70.72	70.90	71.09	70.90	70.72	70.90
3	67.77	72.93	74.40	74.40	74.22	74.03	73.85	73.66
4	71.09	74.03	74.03	74.95	75.14	75.51	75.87	75.32
5	70.90	72.01	73.48	72.56	72.93	72.19	72.01	72.56
6	68.32	69.06	70.17	71.45	71.27	70.53	71.27	71.27
7	73.48	74.03	74.77	75.32	74.95	74.59	74.03	73.85
8	68.88	70.90	71.64	72.56	72.01	71.82	70.90	71.09
9	70.53	72.01	73.66	74.59	74.40	74.40	74.03	74.59
10	68.14	70.72	71.45	71.64	72.38	72.74	72.93	72.56
11	69.24	72.74	73.30	73.30	73.85	73.85	72.93	73.48
12	70.35	72.19	72.56	73.48	73.48	72.93	72.93	73.30
13	68.14	68.69	70.53	72.19	72.56	72.74	72.74	72.56
14	72.01	73.11	74.95	74.77	75.14	74.95	74.77	75.14
15	72.19	73.66	74.40	74.40	73.48	73.30	73.85	73.66
16	70.53	73.30	73.66	74.95	74.59	75.14	75.69	74.77
17	70.17	72.01	71.82	73.85	73.66	73.48	74.03	73.66
18	69.98	71.09	71.45	71.45	71.45	71.64	71.27	71.09
19	67.40	70.35	71.27	71.82	71.27	71.64	71.45	71.27
20	71.45	72.74	73.85	73.66	73.48	73.66	73.11	72.93
Media	69.99	71.92	72.79	73.31	73.27	73.21	73.16	73.07
Desviación estándar	1.62	1.55	1.52	1.38	1.34	1.43	1.54	1.39

4.2. Resultados de clasificación por variedad local utilizando histogramas

Concentrado de resultados de clasificación de 20 ejecuciones, cada ejecución es una selección aleatoria de granos y obtención de características propuestas en este trabajo seguido de la clasificación para la generación de resultados, como se describe en el primer párrafo de éste capítulo.

4.2.1. Resultados con el espacio de color HSI

El siguiente cuadro 4.3 contiene el concentrado de resultados de clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas en 2 y 3 dimensiones del espacio de color HSI.

En los cuadros de resultados se especifica el valor de K que indica la cantidad de vecinos más cercanos que se consideraron en la clasificación para la generación de resultados y con ello tener una comparativa de precisiones con distintos valores para el parámetro K, en el cuadro se tiene los resultados de clasificación utilizando histogramas en 2 y 3 dimensiones.

Los resultados de clasificación por variedad local ocupando HSI son más representativos comparados con los resultados obtenidos con las mediciones puntuales, hay diferencias entre los resultados de clasificación entre histogramas, basado en pruebas estadísticas demuestran no tener diferencias significativas en el uso de histogramas en 2 y 3 dimensiones, computacionalmente es más caro generar histogramas en 3 dimensiones con el espacio de color HSI, por lo tanto, para la identificación de poblaciones por variedad local con el espacio de color HSI resulta adecuado hacer uso de histogramas en 2 dimensiones para la identificación de poblaciones por su variedad local.

4.2.2. Resultados con el espacio de color CIE L*a*b*

En el cuadro 4.4 muestra el concentrado de resultados de 20 ejecuciones de clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 2 y 3 dimensiones del espacio de color CIE L*a*b*.

Los resultados de clasificación muestran diferencias significativas en la media para cada valor del parámetro K, hay una notable diferencia que valida la prueba estadística, es más significativo utilizar los tres canales del espacio de color que solo los canales de cromaticidad para la identificación de variedades locales, a diferencia del espacio de color HSI el espacio de color CIE L*a*b* computacionalmente no es tan caro hacer uso de histogramas en 3 dimensiones y se debe a que el espacio de color CIE L*a*b* mantiene gaussianas definidas por conjuntos de tonos similares, a diferencia en HSI que las frecuencias de tonos están dispersas en el espacio de frecuencias del histograma, lo que se traduce en tener una dispersión de tonos y la búsqueda de la probabilidad conjunta es mayor y requiere más tiempo.

4.2.3. Resultados con el espacio de color HSI y CIE L*a*b*

En el cuadro 4.5 muestra el concentrado de resultados de clasificación de poblaciones nativas de frijol utilizando histogramas de 2 y 3 dimensiones con la unión de los espacios de color HSI y CIE L*a*b*.

Los resultados de clasificación por variedad local que usa la unión de histogramas de los espacios de color HSI y CIE L*a*b* los resultados son similares a los obtenidos con el espacio de color HSI y difieren de CIE L*a*b*, en base a la unión se complementan y la unión de los histogramas de ambos espacios no muestran diferencias significativas en la utilización de histogramas en 2 y 3 dimensiones, resulta más eficiente en tiempo y costo computacional la utilización de la unión de los histogramas en 2 dimensiones.

La Figura 4.1 de manera gráfica muestra los resultados de clasificación de poblaciones por su variedad local utilizando mediciones puntuales con el espectrofotómetro y utilizando toda la distribución de tonos de color con histogramas en 2 y 3 dimensiones de los espacios de color HSI y CIE L*a*b*.

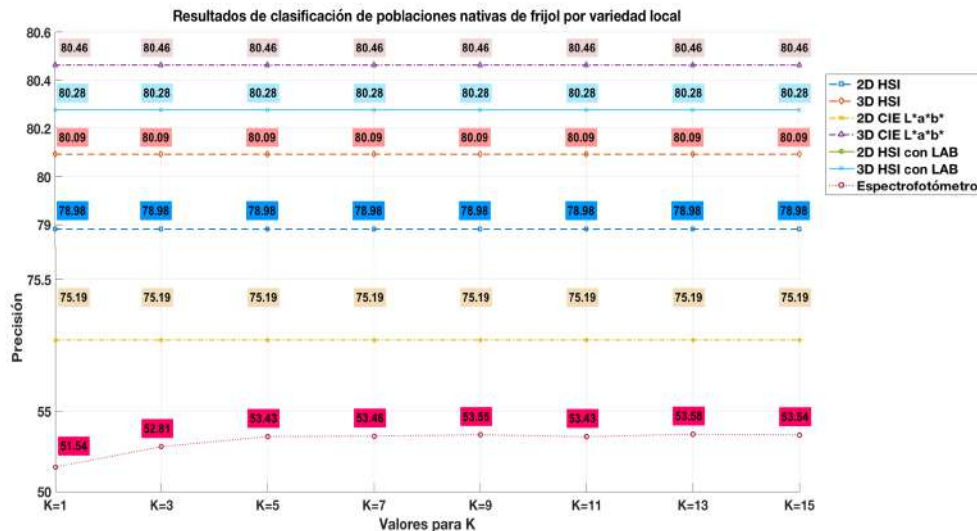


FIGURA 4.1: Gráfica comparativa de clasificación por variedad local

Al observar los resultados de clasificación de poblaciones de acuerdo a su variedad local los porcentajes son iguales para todos los valores de K en histogramas con 2 y 3 dimensiones, esto se debe a que se ocupa K-vecinos más cercanos con distancia ponderada, hace uso del peso que otorga la distancia y no la selección por voto de la clase, si es un K=3, tiene más peso la distancia del vecino más cercano del caso a clasificar, al tener 3 clases distintas cada una tiene su propia distancia ponderada dando siempre prioridad al más cercano, lo mismo sucederá si se pone un K=54, invariablemente el más cercano dominará, por lo tanto el mismo resultado de precisión es igual en todos los valores de K (1, 3, 5, 7, 9, 11, 13, 15), el comportamiento del algoritmo da prioridad a los vecinos locales.

4.3. Resultados de clasificación por color del grano utilizando histogramas

Con la misma determinación de comparar resultados de clasificación utilizando el método aquí propuesto y el método tradicional, se obtuvo resultados de clasificación para la identificación de 54 poblaciones nativas de frijol por color del grano, de igual manera el procedimiento se realizó 20 veces.

4.3.1. Resultados con el espacio de color HSI

El cuadro 4.6 contiene el concentrado de resultados de clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas en 2 y 3 dimensiones del espacio de color HSI.

Los resultados de clasificación por color del grano son más significativos comparados con los obtenidos con las mediciones puntuales realizadas con el espectrofotómetro, la comparativa de resultados entre histogramas de 2 y tres dimensiones para cada valor de K vecinos más cercanos se tiene diferencias significativas, además, los resultados son superiores en la utilización de histogramas de 2 dimensiones y es porque solo se utiliza la parte cromática del espacio de color y resulta más útil para la identificación por color del grano, el componente I de intensidad no aporta mejora significativa, con lo que resulta adecuado emplear solamente la cromaticidad y computacionalmente es menos costoso crear histogramas en 2 dimensiones.

Los resultados de clasificación por color del grano con el espacio de color CIE $L^*a^*b^*$ tiene resultados significativos que mediante pruebas estadísticas es adecuado utilizar los canales a^* y b^* que corresponden a los componentes de cromaticidad por lo que el componente L^* de luminosidad no aporta información suficiente para la identificación que ocupa más vecinos para la clasificación los resultados utilizando 1 y 3 vecinos no presentan diferencias significativas pero por costo computacional es mejor utilizar los canales de cromaticidad y lo que permite tener mejor caracterización.

4.3.2. Resultados con el espacio de color CIE $L^*a^*b^*$

Concentrado de resultados de clasificación de poblaciones nativas de frijol por color del grano para histogramas en 2 y 3 dimensiones del espacio de color CIE $L^*a^*b^*$ Cuadro 4.7.

Los resultados obtenidos de la clasificación para la identificación por color del grano muestran mejores resultados comparados con los obtenidos de las mediciones puntuales del espectrofotómetro, entre histogramas se tienen resultados favorables, con base a la prueba estadística Wilcoxon no hay diferencia significativa para $K=1$ y $K=3$ por cuestiones de costo computacional y por tener resultados más altos es preferible utilizar histogramas en 2 dimensiones que utiliza los canales de cromaticidad, con respecto a K con los demás valores se tienen diferencias significativas y notablemente las precisiones disminuyen más en histogramas de 3 dimensiones, por lo tanto a considerar más vecinos para la clasificación el componente L^* no ofrece mejora.

4.3.3. Resultados con los espacios de color HSI y CIE L*a*b*

Concentrado de resultados de clasificación de poblaciones nativas de frijol por color del grano para histogramas en 2 y 3 dimensiones de la unión de los espacios de color HSI y CIE L*a*b* Cuadro 4.8.

Los resultados de manera similar a los obtenidos donde solo se utiliza un espacio de color para la clasificación de poblaciones por color del granos, se tiene óptimos resultados y la unión puede ser una alternativa más si se desea utilizar dos espacios de color para la clasificación, en base a los resultados de prueba estadística es mejor hacer uso de los canales de cromaticidad por presentar mejor desempeño con histogramas en 2 dimensiones hasta $K=11$ se mantiene arriba del 80% mientras que con histogramas en 3 dimensiones solo hasta $K=7$, con ello queda demostrado que los histogramas en 2 dimensiones resultan adecuados para la clasificación de poblaciones por color del grano.

La Figura 4.2 muestra gráficamente las precisiones de clasificación con mediciones puntuales realizadas con el espectrofotómetro y las precisiones de clasificación con está propuesta.

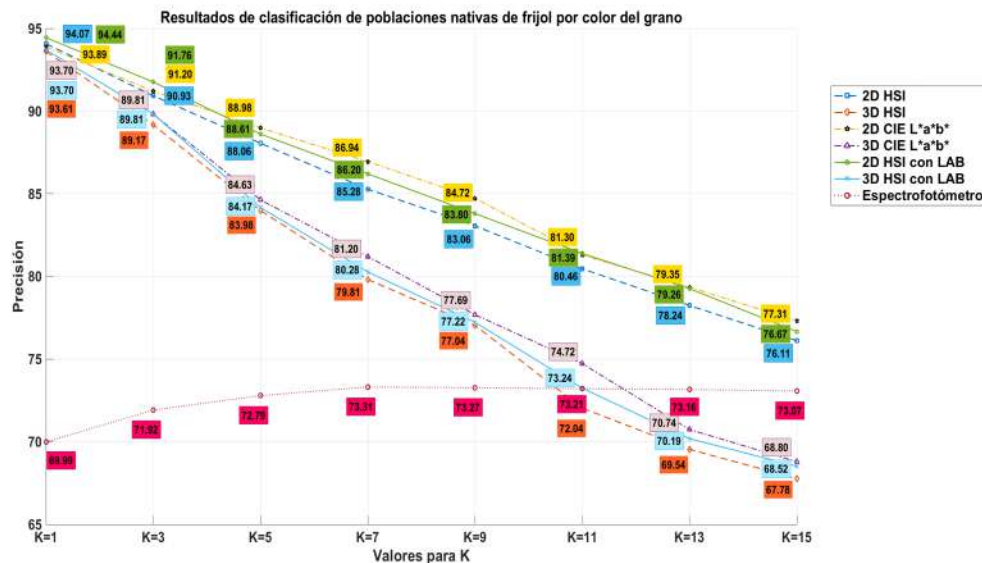


FIGURA 4.2: Gráfica comparativa de clasificación por color del grano

Con los resultados al aplicar prueba estadística Shapiro-Wilk para contrastar la normalidad las precisiones de las 20 ejecuciones de clasificación de acuerdo a la variedad local y clasificación por color del grano, los resultados que provienen de una distribución normal son los generados con histogramas de 2 y 3 dimensiones de los espacios de color HSI y CIE L*a*b*, a excepción de los resultados de clasificación en la que se utiliza la unión de histogramas de 2 dimensiones de los espacios de color HSI con CIE L*a*b* en la que ninguno de sus resultados presenta una distribución uniforme. Para el caso de los resultados de clasificación por color del grano se observó que en todos los vectores existen combinación de distribuciones de los resultados utilizando histogramas de 2 y 3 dimensiones para los dos espacios de color.

Con base a la prueba Wilcoxon para 2 variables no se observó diferencias significativas en la utilización de histogramas de 2 dimensiones con respecto a 3 dimensiones para el espacio de color HSI y la unión de histogramas de 2 dimensiones de HSI con CIE L*a*b*, por lo que cualquiera 2 o 3 dimensiones pueden ser utilizadas para la identificación de poblaciones la única diferencia significativa es para CIE L*a*b* donde es mejor utilizar histogramas en 3 dimensiones para la identificación de las variedades locales con ese espacio de color.

Para la clasificación por color del grano hay diferencias significativas en la utilización de histogramas de 2 a 3 dimensiones, por costo computacional es mejor utilizar histogramas en 2 dimensiones, además, se obtiene resultados con precisiones altos que los ofrecidos utilizando histogramas en 3 dimensiones.

Las diferencias grupales de los resultados aplicando la prueba estadística Kruskal–Wallis con procedimiento de comparación Tukey-Kramer muestran que no hay diferencias en los resultados de clasificación de poblaciones nativas de frijol por variedad local todos presentan la misma media y los resultados de clasificación de poblaciones nativas de frijol y por color del granos todos los grupos presentan diferencias significativas consultar apéndice D.

CUADRO 4.6: Resultados de clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 2 y 3 dimensiones del espacio de color HSI con el algoritmo de clasificación K vecinos más cercanos, donde K significa la cantidad de vecinos utilizados para la clasificación, en el cuadro se incluye los resultados de la aplicación de pruebas estadísticas de normalidad y de significación.

2D								
Ejecuciones	K=1	K=3	K=5	K=7	K=9	K=11	K=13	K=15
1	90.74	87.04	85.19	83.33	79.63	72.22	72.22	66.67
2	96.30	87.04	85.19	81.48	81.48	77.78	72.22	70.37
3	92.59	90.74	87.04	87.04	87.04	79.63	77.78	74.07
4	87.04	87.04	88.89	85.19	81.48	77.78	77.78	77.78
5	94.44	92.59	87.04	85.19	81.48	79.63	79.63	77.78
6	90.74	87.04	87.04	83.33	79.63	77.78	75.93	72.22
7	94.44	92.59	88.89	90.74	85.19	85.19	83.33	81.48
8	98.15	90.74	88.89	87.04	83.33	79.63	79.63	77.78
9	92.59	88.89	85.19	79.63	79.63	79.63	79.63	79.63
10	92.59	88.89	88.89	83.33	81.48	79.63	75.93	72.22
11	92.59	90.74	87.04	81.48	81.48	81.48	79.63	79.63
12	96.30	92.59	92.59	88.89	87.04	85.19	79.63	77.78
13	96.30	88.89	90.74	83.33	79.63	79.63	77.78	75.93
14	92.59	90.74	87.04	85.19	81.48	77.78	75.93	74.07
15	98.15	96.30	88.89	85.19	85.19	83.33	81.48	79.63
16	96.30	96.30	90.74	90.74	88.89	85.19	79.63	77.78
17	92.59	92.59	87.04	85.19	81.48	77.78	75.93	74.07
18	98.15	96.30	88.89	85.19	85.19	83.33	81.48	79.63
19	96.30	96.30	90.74	90.74	88.89	85.19	79.63	77.78
20	92.59	85.19	85.19	83.33	81.48	81.48	79.63	75.93
M	94.07	90.93	88.06	85.28	83.06	80.46	78.24	76.11
DE	2.92	3.50	2.12	3.15	3.08	3.37	2.87	3.70
Shapiro-Wilk								
H	0	0	0	0	1	1	0	0
pValue	0.080416841	0.088916552	0.084894761	0.12357897	0.006946278	0.045076865	0.063516287	0.073073431
3D								
Ejecuciones	K=1	K=3	K=5	K=7	K=9	K=11	K=13	K=15
1	96.30	88.89	83.33	77.78	79.63	74.07	72.22	72.22
2	92.59	88.89	81.48	79.63	75.93	70.37	68.52	64.81
3	94.44	87.04	83.33	83.33	79.63	75.93	72.22	70.37
4	96.30	88.89	85.19	85.19	79.63	77.78	74.07	70.37
5	94.44	88.89	85.19	79.63	75.93	74.07	68.52	66.67
6	92.59	88.89	85.19	81.48	79.63	72.22	72.22	70.37
7	92.59	87.04	83.33	83.33	74.07	68.52	68.52	66.67
8	92.59	92.59	87.04	79.63	77.78	72.22	70.37	68.52
9	94.44	88.89	83.33	74.07	74.07	68.52	66.67	64.81
10	90.74	90.74	83.33	83.33	79.63	66.67	62.96	62.96
11	94.44	88.89	85.19	81.48	77.78	74.07	74.07	68.52
12	90.74	90.74	83.33	79.63	77.78	70.37	68.52	68.52
13	94.44	88.89	83.33	79.63	79.63	74.07	72.22	72.22
14	96.30	88.89	83.33	75.93	72.22	72.22	70.37	68.52
15	94.44	88.89	83.33	79.63	77.78	72.22	68.52	68.52
16	88.89	88.89	81.48	75.93	75.93	66.67	64.81	62.96
17	94.44	88.89	85.19	81.48	77.78	74.07	68.52	64.81
18	88.89	85.19	81.48	77.78	72.22	68.52	66.67	64.81
19	96.30	90.74	81.48	72.22	74.07	70.37	68.52	68.52
20	96.30	92.59	90.74	85.19	79.63	77.78	72.22	70.37
Media	93.61	89.17	83.98	79.81	77.04	72.04	69.54	67.78
Desviación estándar	2.36	1.73	2.19	3.50	2.58	3.29	2.97	2.84
Shapiro-Wilk								
H	1	1	1	0	1	0	0	0
pValue	0.017001213	0.004180295	0.001768534	0.445419797	0.007326901	0.365512809	0.184783331	0.139611979
Wilcoxon								
p	0.617617238	0.110739829	5.66259E-06	3.17008E-05	3.40494E-07	4.98972E-07	1.93415E-07	8.98821E-07
h	0	0	1	1	1	1	1	1

CUADRO 4.8: Resultados de clasificación por color del grano utilizando histogramas de 2 y 3 dimensiones de los espacios de color HSI y CIE L*a*b* con el algoritmo de clasificación K vecinos más cercanos, donde K significa la cantidad de vecinos utilizados para la clasificación, en el cuadro se incluye los resultados de la aplicación de pruebas estadísticas de normalidad y de significación.

2D								
Ejecuciones	K=1	K=3	K=5	K=7	K=9	K=11	K=13	K=15
1	90.74	87.04	85.19	81.48	74.07	72.22	70.37	66.67
2	94.44	85.19	85.19	83.33	83.33	79.63	75.93	70.37
3	92.59	90.74	87.04	88.89	85.19	81.48	75.93	75.93
4	90.74	90.74	88.89	88.89	83.33	77.78	77.78	77.78
5	94.44	92.59	88.89	85.19	83.33	81.48	81.48	79.63
6	88.89	87.04	87.04	83.33	83.33	81.48	79.63	74.07
7	94.44	92.59	88.89	87.04	87.04	85.19	81.48	79.63
8	96.30	90.74	88.89	87.04	83.33	79.63	79.63	77.78
9	92.59	88.89	85.19	81.48	81.48	77.78	75.93	77.78
10	92.59	90.74	88.89	88.89	83.33	81.48	77.78	74.07
11	94.44	92.59	87.04	85.19	85.19	83.33	79.63	79.63
12	94.44	92.59	90.74	87.04	83.33	83.33	81.48	79.63
13	98.15	92.59	88.89	85.19	83.33	77.78	75.93	75.93
14	92.59	88.89	87.04	85.19	79.63	75.93	75.93	72.22
15	98.15	96.30	90.74	85.19	85.19	83.33	81.48	77.78
16	98.15	98.15	92.59	90.74	88.89	88.89	87.04	81.48
17	94.44	92.59	88.89	87.04	81.48	81.48	77.78	75.93
18	98.15	96.30	90.74	85.19	85.19	83.33	81.48	77.78
19	98.15	98.15	92.59	90.74	88.89	88.89	87.04	81.48
20	94.44	90.74	88.89	87.04	87.04	83.33	81.48	77.78
Media	94.44	91.76	88.61	86.20	83.80	81.39	79.26	76.67
Desviación estándar	2.75	3.53	2.19	2.65	3.28	4.01	3.92	3.72
Shapiro-Wilk								
H	0	0	0	0	1	0	0	1
pValue	0.058756259	0.251949668	0.089906993	0.299406323	0.011908288	0.269760356	0.076601195	0.034134002
3D								
Ejecuciones	K=1	K=3	K=5	K=7	K=9	K=11	K=13	K=15
1	96.30	88.89	85.19	79.63	81.48	75.93	74.07	72.22
2	92.59	90.74	81.48	79.63	77.78	72.22	68.52	68.52
3	94.44	88.89	83.33	85.19	79.63	75.93	72.22	68.52
4	96.30	88.89	85.19	85.19	77.78	77.78	75.93	70.37
5	94.44	88.89	85.19	77.78	75.93	75.93	68.52	66.67
6	92.59	88.89	85.19	81.48	79.63	74.07	72.22	70.37
7	92.59	87.04	83.33	77.78	74.07	68.52	66.67	66.67
8	92.59	90.74	87.04	79.63	77.78	74.07	70.37	70.37
9	94.44	90.74	85.19	77.78	75.93	74.07	64.81	64.81
10	90.74	90.74	83.33	83.33	77.78	68.52	64.81	62.96
11	94.44	90.74	85.19	83.33	79.63	75.93	75.93	72.22
12	92.59	90.74	85.19	81.48	77.78	74.07	70.37	68.52
13	94.44	90.74	83.33	81.48	79.63	74.07	72.22	72.22
14	96.30	88.89	83.33	77.78	74.07	72.22	70.37	70.37
15	94.44	90.74	83.33	79.63	77.78	74.07	70.37	68.52
16	88.89	87.04	81.48	77.78	75.93	66.67	64.81	64.81
17	94.44	90.74	85.19	81.48	77.78	75.93	70.37	66.67
18	88.89	87.04	81.48	77.78	72.22	68.52	66.67	64.81
19	96.30	90.74	81.48	72.22	72.22	70.37	68.52	68.52
20	96.30	94.44	88.89	85.19	79.63	75.93	75.93	72.22
M	93.70	89.81	84.17	80.28	77.22	73.24	70.19	68.52
DE	2.28	1.75	1.94	3.25	2.56	3.15	3.60	2.82
Shapiro-Wilk								
H	1	1	1	0	0	1	0	0
pValue	0.012094851	0.003254015	0.027521728	0.081049937	0.075473521	0.018688869	0.177155838	0.13834261
Wilcoxon								
p	0.47684003	0.028028021	2.23458E-06	4.04559E-06	1.16954E-06	9.39658E-07	6.4276E-07	1.46328E-06
h	0	1	1	1	1	1	1	1

Capítulo 5

Discusión

La diferencia entre una variedad de frijol y una población de granos de frijol es que la primera todos los granos reúnen similitud en tamaño color, forma, textura entre otras características que serán de una mínima variación y las poblaciones son variedades que pueden estar conformadas por granos similares o combinación de granos de distintos colores, los trabajos reportan estudio sobre variedades de frijol y las mediciones de color toman promedios y al ser variedades con granos del mismo tipo registran un promedio, lo que no resulta útil si tenemos poblaciones con granos de distintos colores.

Actualmente algunos trabajos de la literatura reportan el uso de sistemas de visión con iluminación directa sobre los granos que deja manchas de brillo sobre su superficie y las imágenes no presentan la nitidez requerida para caracterizar por color o utilizan un escáner para la adquisición de imágenes de los granos teniendo reflejos y aplicando filtro pasa bajas para el suavizado de la imagen y difuminar sombras, al utilizar solo un promedio no dan importancia a la calidad de la imagen porque utilizan otras propiedades físicas para caracterizar, el presente trabajo ofrece una mejora a la iluminación del sistema de visión que es un prototipo de iluminación con caja de luz translúcida que difumina luz, reduce sombras y brillo en granos, además, con la adecuada distancia para enfoque y configuración de la cámara fotográfica digital mejora la nitidez que hace posible capturar las estrías, moteado y todos los pigmentos de color de la superficie del grano.

Los espacio de color HSI y CIE L*a*b* resultaron ser apropiados para la descripción del color y que a diferencia del color RGB tienen separada la cromaticidad de la luminosidad que permite lograr caracterización de poblaciones nativas de frijol, tanto para la identificación por variedad local como por coloración del grano.

El algoritmo de segmentación crecimiento de región por color es otra alternativa viable para la separación de los granos con el fondo, en la literatura se hacen propuestas interesantes de segmentación, para este estudio y a manera de contribución se demuestra que existe otra alternativa útil para segmentar y que puede ser propicia para utilizar varios fondos que permitan contrastar con los objetos de interés de estudio.

Otra contribución es la utilización de histogramas formados con la probabilidad conjunta de los valores de los píxeles de los canales de cromaticidad y de luminosidad, para representar toda la distribución de tonos de color de los granos de frijol.

La paleta de color creada para dar descripción a cada una de las 54 poblaciones de este estudio

fue preciso crearla con los granos de las poblaciones, existen paletas de colores definidas con el inconveniente que son solo para igualación de colores y cada categoría solo considera una pequeña muestra de colores planos.

Con variabilidad de color en los granos de las poblaciones y para la correcta identificación se dividió a cada población en dos conjuntos para mantener la homogeneidad de las poblaciones y que permitiera su identificación de forma semejante a la identificación que se hace con la medición de color en los granos de las poblaciones de manera tradicional, este método de separación resulto de gran utilidad para mantener homogeneidad en poblaciones, existen casos de selección aleatoria con predominancia de granos de un color en un conjunto que en otro, el sesgo hacia el color predominante da el comportamiento que debe tener sobre la identificación de poblaciones clasificando hacia las poblaciones del color dominante, en base a lo anterior podría considerarse interrogantes de investigación como: ¿Por qué el dominante de una especie de granos sobre otra de la misma población en un cultivo?, ¿Por qué hay diferencia de tamaños en los granos de poblaciones similares?, ¿Qué poblaciones de frijol sembrar en determinados periodos de año?, ¿Qué variedades son más tolerantes a ciertas condiciones climáticas?.

Para la identificación de poblaciones por variedad local es más conveniente utilizar histogramas de 3 dimensiones cuando se hace uso del espacio de color CIE $L^*a^*b^*$ ya que en este caso L^* aporta mejora en la identificación, para el caso de HSI no hay diferencia entre la utilización de 2 y 3 dimensiones por lo que puede resultar más práctico utilizar solo los canales de cromaticidad.

Las mediciones puntuales realizadas con el espectrofotómetro dan como resultado de 51.54 %, 52.81 % para K con 1, 3 respectivamente, estos porcentajes son inferiores comparados con los porcentajes resultantes de este estudio, los porcentajes ocupando histogramas de 2 dimensiones del espacio de color HSI se obtuvo 78.98 % de precisión y para el espacio de color CIE $L^*a^*b^*$ se obtuvo el 75.19 % para los mismos valores del parámetro K que se utilizaron para las mediciones puntuales, los resultados de clasificación ocupando histogramas de 3 dimensiones se obtuvieron los resultados de HSI con 80.09 % para identificación de variedades locales y de 80.46 % para CIE $L^*a^*b^*$ en todos los valores de K.

Para la clasificación por color del grano para los dos primeros valores de K con 1, 3 se tiene que para la mediciones puntuales los porcentajes de 69.99 %, 71.92 % que resultan ser inferiores a los porcentajes de nuestra propuesta y solo se estancan a medida que se especifican más vecinos, los porcentajes obtenidos para los histogramas en 2 dimensiones en la identificación por color del grano para HSI es 94.07 %, 90.93 % de precisión y para el espacio de color CIE $L^*a^*b^*$ 93.89 %, 91.20 % y para la clasificación con histogramas en 3 dimensiones con resultados de HSI en 93.61 %, 89.17 % de precisión y para CIE $L^*a^*b^*$ 93.70 %, 89.81 % de precisión, estos resultados muestran que éste método al hacer uso de toda la distribución de color representada en histogramas es adecuado para la correcta clasificación de poblaciones nativas de frijol.

Trabajos relacionados que obtienen resultados de identificación altos utilizan promedio de color y complementan con otras características como la morfológica o de textura y resultan adecuados para sus enfoques de diseño, baso en nuestro enfoque quedan limitados al no considerar la distribución de tonos de color que puede presentar un grano y que unido a este problema las poblaciones están conformadas por granos de distintos colores.

Para la identificación de poblaciones nativas de frijol por color del grano resulta apropiado utilizar histogramas de 2 dimensiones formados por los canales de cromaticidad que son los canales H y S del espacio de color HSI y de CIE L*a*b* los canales a* y b*, el canal de intensidad I y luminosidad L* no aportan mejora solo incrementa el costo computacional.

Identificar las propiedades colorimétricas de 54 poblaciones nativas de frijol mediante la distribución de color de los granos de cada población utilizando histogramas para la representación de toda la distribución con los espacios de color HSI y CIE L*a*b* para caracterizar el color de las poblaciones, los resultados muestran que este método es mejor en caracterización de poblaciones nativas de frijol al considerar toda la distribución de tonos de color, el espectrofotómetro es un instrumento de medición que ofrece valores puntuales de estimación del color y poco práctico para obtener la gama de colores con ese método, y a pesar de que existen trabajos de sistema de visión por computadora relativos a la clasificación de variedades de frijol consideran un promedio de color y se apoyan de otras características como las morfológica y de textura para identificar variedades, aunque esos trabajos están diseñados para cumplir solo el objetivo para el que fueron diseñados y ante la integración de nuevas variedades sus procedimientos metodológico requieren nuevas mejoras para la correcta identificación y no cubren la necesidad de caracterización utilizando únicamente las propiedades colorimétricas. Este trabajo ofrece la metodología idónea para la caracterización de poblaciones nativas de frijol y no requiere hacer cambios a los algoritmos para considerar nuevas variedades que se requieran integrar.

La metodología propuesta utilizada para la clasificación de poblaciones nativas de frijol, no queda limitada y puede ser utilizada para la clasificación de frutos y hortalizas en las que se requiera la evaluación por color para la identificación de algún estado de madurez, de oxidación entre otras necesidades relativas a la coloración, al utilizar solo la distribución de color resulta un método flexible para incorporarse en otros estudios de interés.

Capítulo 6

Conclusiones y trabajo futuro

Se confirma la hipótesis que afirma que la utilización de toda la distribución de color de los granos de frijol es posible tener una mejor identificación y clasificación de las 54 poblaciones nativas de frijol.

Los histogramas como descriptores de color resultaron apropiados para representar toda la distribución de color de los granos de las poblaciones, además los espacios de colores HSI y CIE $L^*a^*b^*$ describen y caracterizan el color de las poblaciones.

Como aportación de este trabajo ofrece la identificación de poblaciones nativas de frijol con apoyo de un prototipo de iluminación y una cámara fotográfica para la adquisición de imágenes, es vital un entorno de constante iluminación que evite luz de otras fuentes de distintas intensidades, de manera similar a la visión humana que requiere una adecuada iluminación que no sea afectada por otras condiciones.

El algoritmo de clasificación es una herramienta imprescindible para clasificar, la utilización de K vecinos cercanos con distancia ponderada cumplió el compromiso de identificación y clasificación de poblaciones nativas de frijol utilizando histogramas.

La elaboración de un método automatizado que permita identificar poblaciones nativas de frijol como una alternativa fiable que ayude a clasificar poblaciones, aportando sugerencias como una segunda opinión para el experto y que le brinde la fiabilidad de identificación, además, está solución no requiera de personal capacitado.

Como conclusión podemos afirmar que la propiedad física del color de los granos es un buen elemento para identificar y clasificar poblaciones nativas de frijol, haciendo uso de toda la distribución de tonos de color que conforma al objeto u objetos de interés para su caracterización, en el presente trabajo se demostró que las mediciones puntuales realizadas con un espectrofotómetro no es robusta para la identificación de poblaciones nativas de frijol y a pesar de que existen trabajos para identificación de variedades de frijol ocupan un promedio del color, la utilización de toda la distribución de tonos de color resulta adecuado para la caracterización de las poblaciones nativas de frijol.

El método propuesto en este estudio es un método económico comparado con el precio de un espectrofotómetro para sólidos por tratarse de un instrumento de laboratorio, la metodología no está limitada, es flexible para la incorporación de nuevas poblaciones nativas de frijol al conjunto de entrenamiento para hacerlo más robusto ante la existencia de poblaciones nativas de frijol de

otros colores, también es un método que no está solo restringido a las poblaciones nativas de frijol y que puede ser utilizado en otros estudios relacionados con el color.

Este método es no destructivo al evitar la destrucción del grano evitando todo el procedimiento químico de laboratorio que requiere para estimar composición química logrando la reducción del tiempo y no requerirá de un experto para su utilización.

Como trabajo futuro para este proyecto basado en este enfoque se propone investigar la viabilidad de estimación de antocianinas con base a la similitud de color entre las poblaciones nativas de frijol y otras propiedades químicas, ya que existen estudios que han investigado la relación entre el color y la composición química[10]. También, como trabajo futuro esta metodología puede ser empleada para análisis de frutos y hortalizas utilizando las propiedades colorimétricas.

Se propone como trabajo futuro explorar otros algoritmos de clasificación que utilicen las probabilidades calculadas de este trabajo como lo puede ser el clasificador bayesiano o algunos otros clasificadores para tener una comparativa de desempeño entre clasificadores.

Apéndice A

Georreferencia de las 54 poblaciones nativas de frijol

CUADRO A.1: Georreferencia de las 54 poblaciones nativas de frijol

Población	Municipio	Localidad	Latitud	Longitud	Altitud
POB-02	Miahuatlaán de Porfirio Díaz	Agua de Sol	16°28'24"	96°43'55"	1505
POB-03	Taniche	Teniche	16°33'57"	96°45'11"	1426
POB-04	Taniche	Teniche	16°33'57"	96°45'11"	1426
POB-05	San Agustín Amatengo	San Agustín Amatengo	16°30'37"	96°47'21"	1361
POB-06	San Agustín Amatengo	San Agustín Amatengo	16°30'37"	96°47'21"	1361
POB-08	Magdalena Jaltepec	Morelos	17°17'23"	97°16'59"	2055
POB-09	Ánimas Trujano	Ánimas Trujano	16°59'18"	96°42'46"	1518
POB-11	San Juan Tamazola	Yucuxina	17°12'18"	97°11'54"	2083
POB-13	San Juan Tamazola	Yucuxina	17°12'18"	97°11'54"	2083
POB-17	Magdalena Jaltepec	San Miguel	17°18'34"	97°14'49"	2040
POB-18	Magdalena Jaltepec	San Miguel	17°18'34"	97°14'49"	2040
POB-19	Santo Domingo Yanhuitlán	Tooxi	17°33'25"	97°22'56"	2406
POB-20	Santiago Tlazoyaltepec	Cañada de Yuxcimil	17°0'31"	96°58'34"	2477
POB-22	Santiago Tlazoyaltepec	Cañada de Yuxcimil	17°0'31"	96°58'34"	2477
POB-23	Santiago Tlazoyaltepec	Río de San Juan	17°1'7"	96°59'25"	2225
POB-26	Santiago Tlazoyaltepec	El Gachupín	17°0'28"	96°57'51"	2681

Continúa en la siguiente página

Cuadro A.1 – Georreferencia de las 54 poblaciones nativas de frijol

Población	Municipio	Localidad	Latitud	Longitud	Altitud
POB-28	Santiago Tlazoyaltepec	La Unión	17°2'15"	96°59'44"	2458
POB-29	Ixtlán de Juárez	Santa María Yahuiche	17°18'0"	96°28'53"	1836
POB-30	Santiago Laxopa	Santiago Laxopa	17°13'1"	96°18'40"	1942
POB-31	Santiago Laxopa	Santiago Laxopa	17°13'1"	96°18'40"	1942
POB-32	San Juan Evangelista	San Juan Evangelista Analco	17°24'24"	96°32'20"	2065
POB-33	San Juan Evangelista	San Juan Evangelista Analco	17°24'24"	96°32'20"	2065
POB-34	Santiago Xiacuí	San Andrés Yatuni	17°15'6"	96°24'11"	2337
POB-35	Santiago Laxopa	Santiago Laxopa	17°13'1"	96°18'40"	1942
POB-39	Reyes Etna	Reyes Etna	17°12'18"	96°48'37"	1647
POB-40	San Juan Bautista Guelache	San Gabriel Etna	17°12'58"	96°46'0"	1822
POB-41	San Juan Bautista Guelache	San Gabriel Etna	17°12'58"	96°46'0"	1822
POB-43	Santa María Tlahuitoltepec	Santa María Tlahuitoltepec	17°5'35"	96°3'52"	2239
POB-44	Santa María Tlahuitoltepec	Santa María Tlahuitoltepec	17°5'35"	96°3'52"	2239
POB-46	San Pedro y San Pablo Ayutla	San Pedro y San Pedro Ayutla	17°1'32"	96°4'48"	2040
POB-48	San Simón Almolongas	La Estancia	16°26'4"	96°41'26"	1516
POB-49	Heroica Ciudad de Ejutla de Crespo	Barranca Larga	16°29'30"	96°44'10"	1528
POB-50	Miahuatlán de Porfitio Díaz	San Pedro Amatlán	16°19'39"	96°26'57"	1566
POB-51	Santa Cruz Xitla	Santa Cruz Xitla	16°19'17"	96°40'20"	1804
POB-52	Santa Cruz Xitla	Santa Cruz Xitla	16°19'17"	96°40'20"	1804
POB-54	Santa Catarina Cuixtla	Santa Catarina Cuitla	16°18'16"	96°38'43"	1791
POB-55	Santa Catarina Cuixtla	Santa Catarina Cuitla	16°18'16"	96°38'43"	1791
POB-56	Miahuatlán de Profirio Díaz	San Miguel Yogovana	16°16'53"	96°37'58"	1758
POB-57	San Agustín Amatengo	San Agustín Amatengo	16°30'37"	96°47'21"	1361

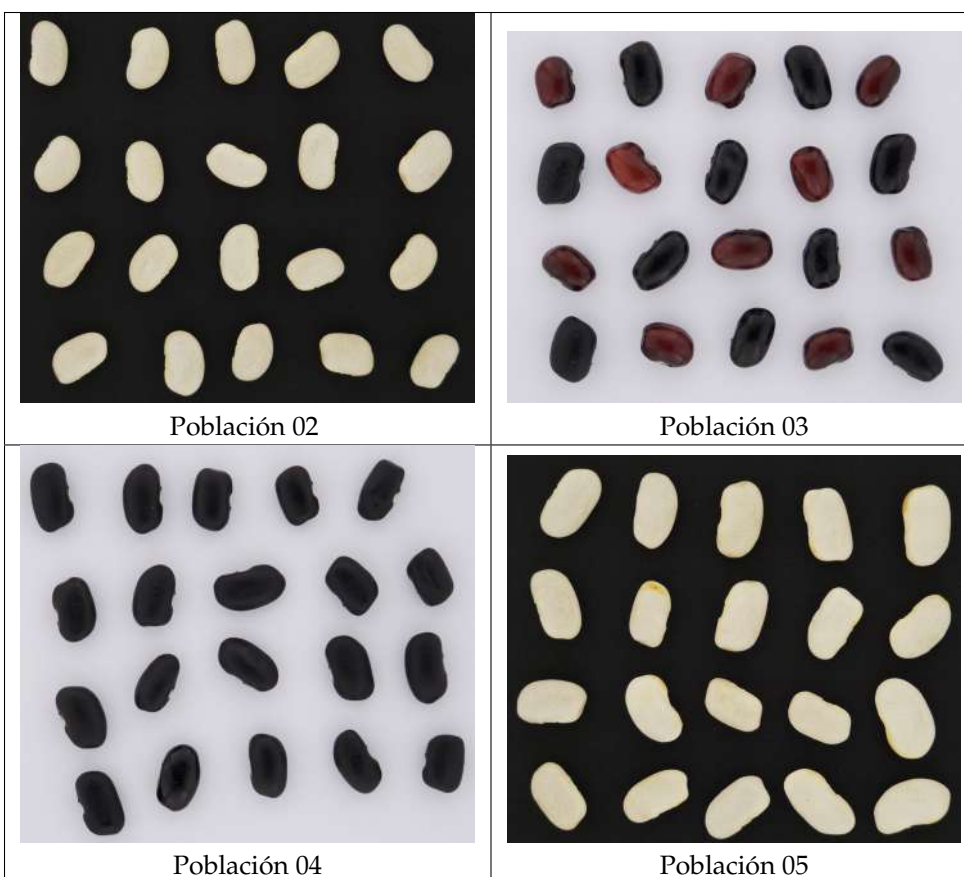
Continúa en la siguiente página

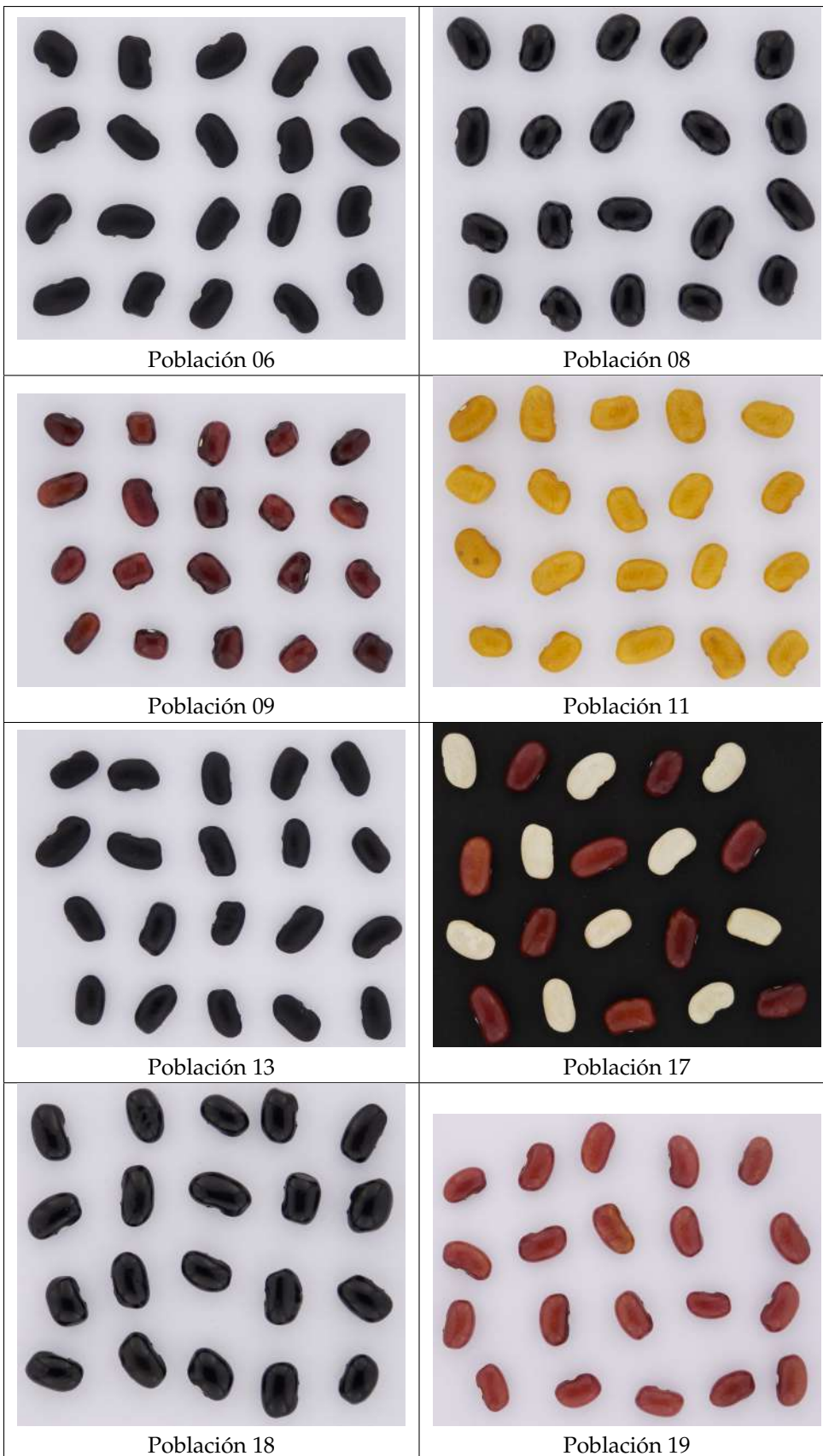
Cuadro A.1 – Georreferencia de las 54 poblaciones nativas de frijol

Población	Municipio	Localidad	Latitud	Longitud	Altitud
POB-58	San Juan Tamazola	Yucuxina	17°12'28"	97°11'54"	2083
POB-59	San Juan Tamazola	Yucuxina	17°12'28"	97°11'54"	2083
POB-61	San Juan Tamazola	Yucuxina	17°12'28"	97°11'54"	2083
POB-62	Tamazulápan del Progreso	Río del Oro	17°41'41"	97°36'41"	1894
POB-63	Villa Tejúpam de la Unión	Yodobada	17°37'16"	97°26'18"	2247
POB-64	Silacayoápam	Santiago Asunción	17°31'5"	98°1'47"	1921
POB-65	Silacayoápam	Santiago Asunción	17°31'5"	98°1'47"	1921
POB-66	Silacayoápam	Santiago Asunción	17°31'5"	98°1'47"	1921
POB-67	Silacayoápam	Santiago Asunción	17°31'5"	98°1'47"	1921
POB-68	Silacayoápam	Santiago Asunción	17°31'5"	98°1'47"	1921
POB-69	Santiago Juxtlahuaca	Santa María Yucunicoco	17°14'15"	97°55'8"	2458
POB-71	Santiago Juxtlahuaca	Santo Domingo del Progreso	17°17'57"	97°55'51"	2684
POB-72	Santiago Juxtlahuaca	Tinuma de Zaragoza	17°20'34"	97°54'26"	2356
POB-74	Santiago Juxtlahuaca	Santa María Yucunicoco	17°14'15"	97°55'8"	2458
POB-76	San Sebastián Tecomaxtlahuaca	Santa Cruz Rancho Viejo	17°23'51"	98°3'8"	1869

Apéndice B

54 Poblaciones nativas de frijol







Población 20



Población 22



Población 23



Población 26



Población 28



Población 29



Población 30



Población 31



Población 32



Población 33



Población 34



Población 35



Población 39



Población 40



Población 41



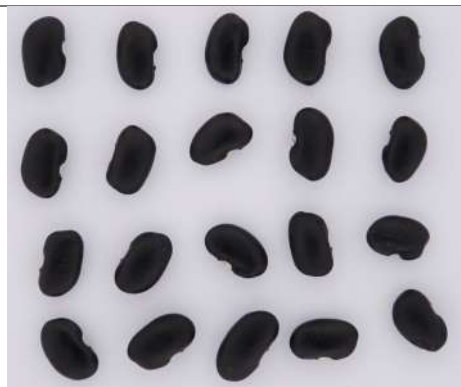
Población 43



Población 44



Población 46



Población 48



Población 49



Población 50



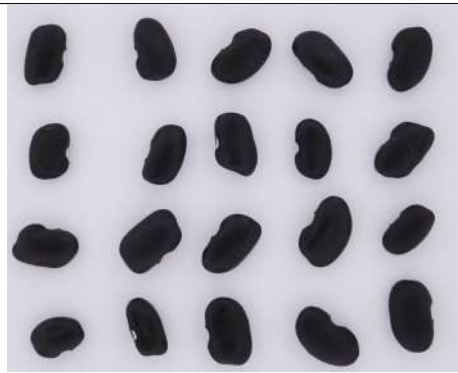
Población 51



Población 52



Población 54



Población 55



Población 56



Población 57



Población 58



Población 59



Población 61



Población 62



Población 63



Población 64



Población 65



Población 66



Población 67



Población 68



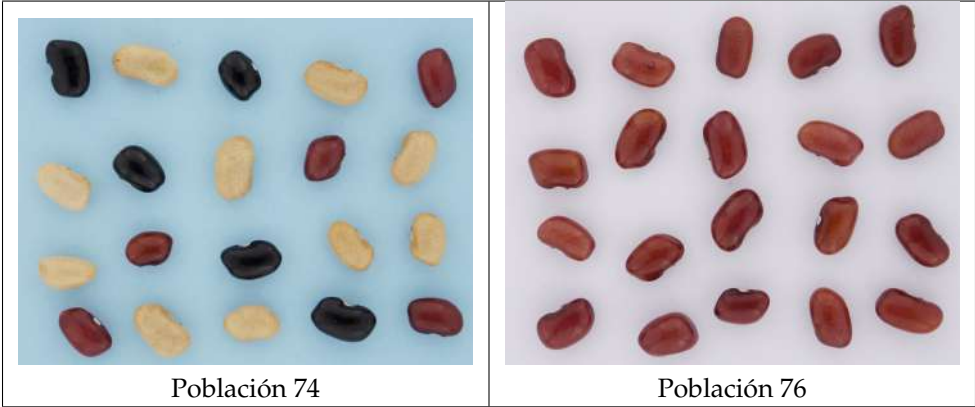
Población 69



Población 71



Población 72



Apéndice C

Descripción de color para las 54 poblaciones de frijol

CUADRO C.1: Taxonomía de color.

Población	Descripción	Identificador
Pob-02	Blanco	3
Pob-03	Mezcla negro y rojo	17
Pob-04	Negro	19
Pob-05	Blanco	3
Pob-06	Negro	19
Pob-08	Negro	19
Pob-09	Rojo	20
Pob-11	Amarillo	1
Pob-13	Negro	19
Pob-17	Mezcla blanco y rojo	12
Pob-18	Negro	19
Pob-19	Rojo	20
Pob-20	Mezcla amarillo, morado, morado con negro, negro y rosa	8
Pob-22	Mezcla morado, morado con negro, rosa y rosa con negro	16
Pob-23	Mezcla amarillo con negro, café, negro, rojo y rosa con negro	4
Pob-26	Mezcla amarillo, negro y rosa	10
Pob-28	Mezcla negro y rojo	17
Pob-29	Negro	19
Pob-30	Negro	19
Pob-31	Mezcla morado y negro	15
Pob-32	Mezcla amarillo, amarillo con café y negro	7
Pob-33	Mezcla amarillo, negro y rosa	10
Pob-34	Mezcla amarillo y amarillo con negro	5
Pob-35	Mezcla morado y negro	15
Pob-39	Mezcla café y rojo	14
Pob-40	Rojo	20
Pob-41	Negro	19
Pob-43	Mezcla amarillo y rojo	6
Continúa en la siguiente página →		

Cuadro C.1 – ↔ Complementa la página anterior

Población	Descripción	Identificador
Pob-44	Negro	19
Pob-46	Negro	19
Pob-48	Negro	19
Pob-49	Negro	19
Pob-50	Blanco	3
Pob-51	Mezcla amarillo, negro y rojo	9
Pob-52	Negro	19
Pob-54	Mezcla negro, rojo y rosa	18
Pob-55	Negro	19
Pob-56	Mezcla blanco y negro	11
Pob-57	Amarillo	1
Pob-58	Rojo	20
Pob-59	Amarillo	1
Pob-61	Negro	19
Pob-62	Rojo	20
Pob-63	Rojo	20
Pob-64	Rojo	20
Pob-65	Amarillo	1
Pob-66	Mezcla negro y rojo	17
Pob-67	Blanco	3
Pob-68	Amarillo con negro	2
Pob-69	Mezcla morado y negro	15
Pob-71	Rosa con café	21
Pob-72	Mezcla amarillo, negro y rojo	9
Pob-74	Mezcla blanco, negro y rojo	13
Pob-76	Rojo	20

Apéndice D

Resultado de prueba estadística Kruskal–Wallis

D.1. Resultados gráficos estadísticos de clasificación de poblaciones nativas de frijol por su variedad local

D.1.1. Utilizando histogramas de 2 dimensiones del espacio de color HSI

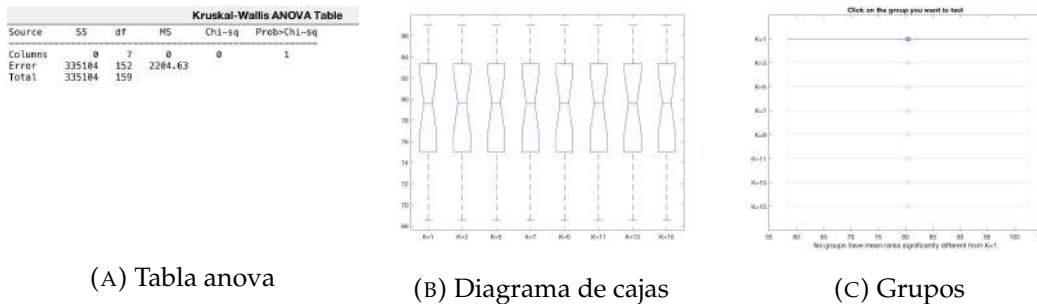


FIGURA D.1: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 2 dimensiones del espacio de color HSI

D.1.2. Utilizando histogramas de 3 dimensiones del espacio de color HSI

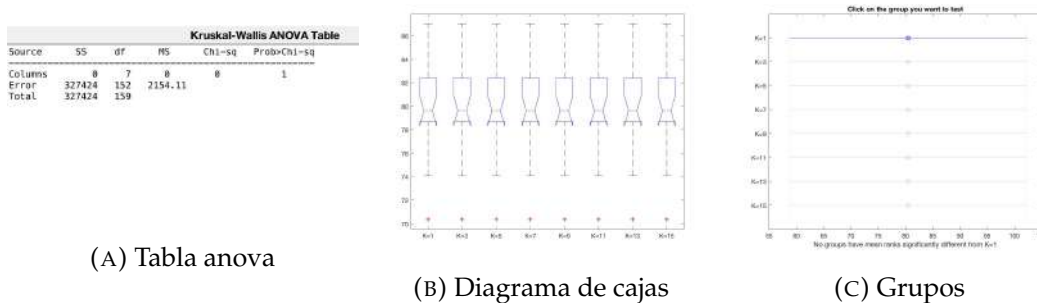


FIGURA D.2: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 3 dimensiones del espacio de color HSI

D.1.3. Utilizando histogramas de 2 dimensiones del espacio de color CIE L*a*b*

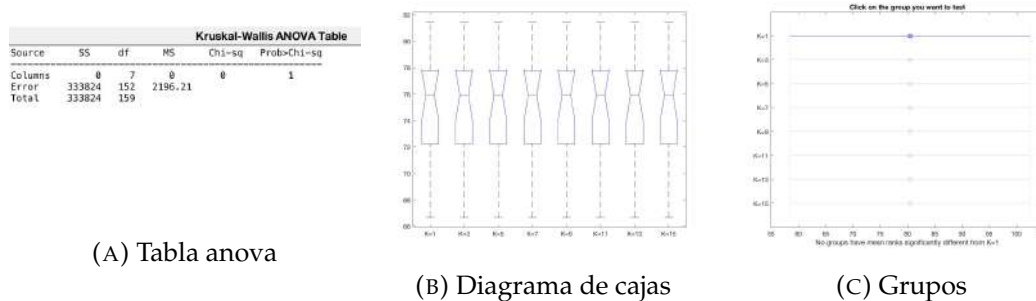


FIGURA D.3: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 2 dimensiones del espacio de color CIE L*a*b*

D.1.4. Utilizando histogramas de 3 dimensiones del espacio de color CIE L*a*b*

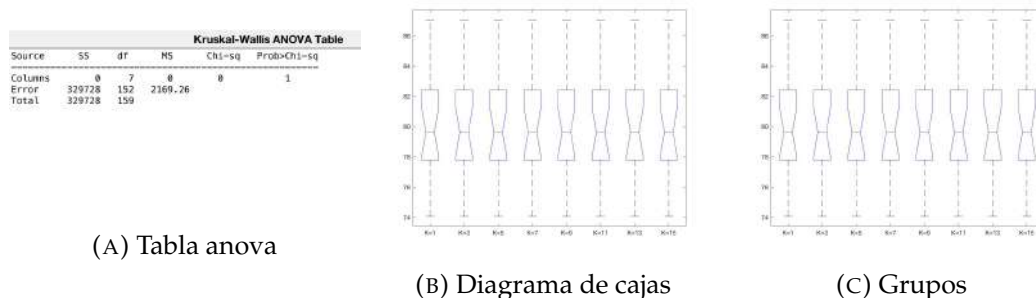


FIGURA D.4: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 3 dimensiones del espacio de color CIE L*a*b*

D.1.5. Utilizando la unión de los histogramas de 2 dimensiones de los espacios de color HSI con CIE L*a*b*

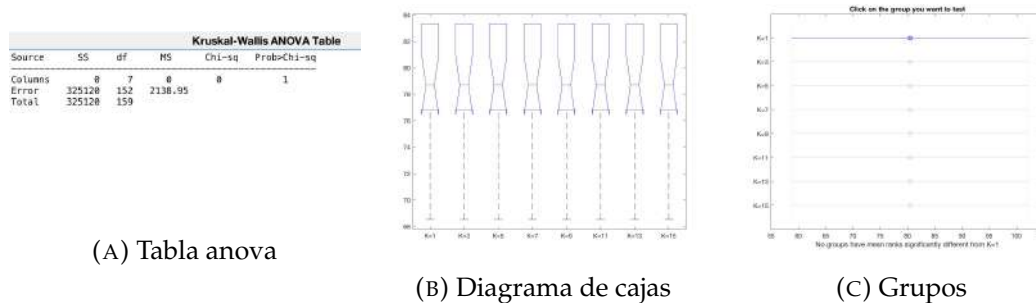


FIGURA D.5: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 2 dimensiones de los espacios de color HSI con CIE L*a*b*

D.1.6. Utilizando la unión de los histogramas de 3 dimensiones de los espacios de color HSI con CIE L*a*b*

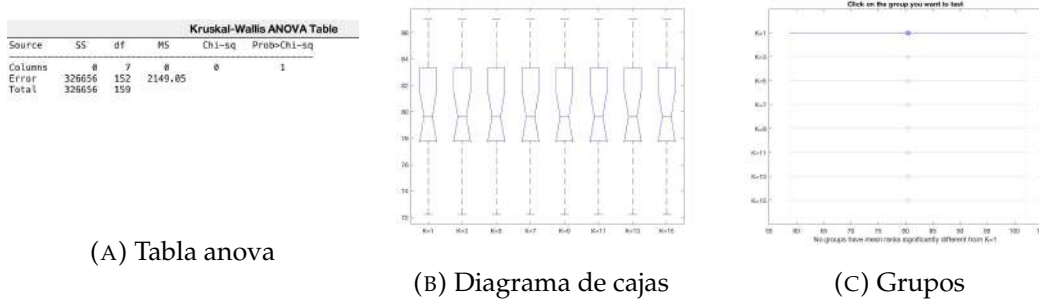


FIGURA D.6: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por su variedad local utilizando histogramas de 3 dimensiones de los espacios de color HSI con CIE L*a*b*

D.2. Resultados gráficos estadísticos de clasificación de poblaciones nativas de frijol por color del grano

D.2.1. Utilizando histogramas de 2 dimensiones del espacio de color HSI

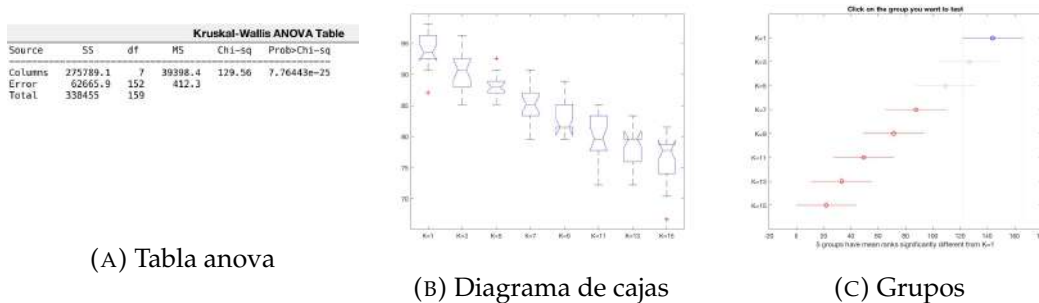


FIGURA D.7: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 2 dimensiones del espacio de color HSI

D.2.2. Utilizando histogramas de 3 dimensiones del espacio de color HSI

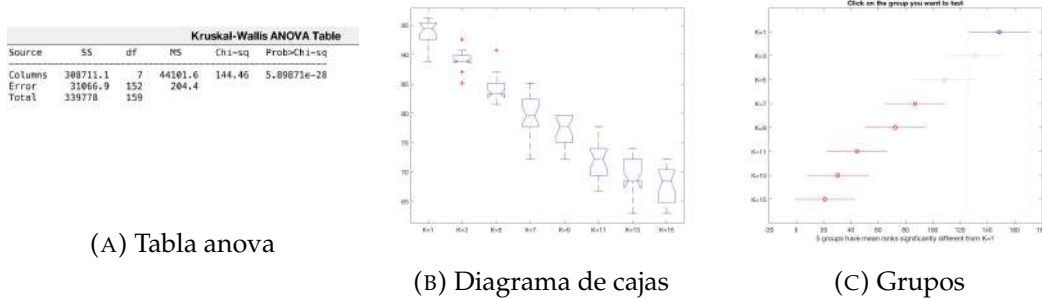


FIGURA D.8: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 3 dimensiones del espacio de color HSI

D.2.3. Utilizando histogramas de 2 dimensiones del espacio de color CIE L*a*b*

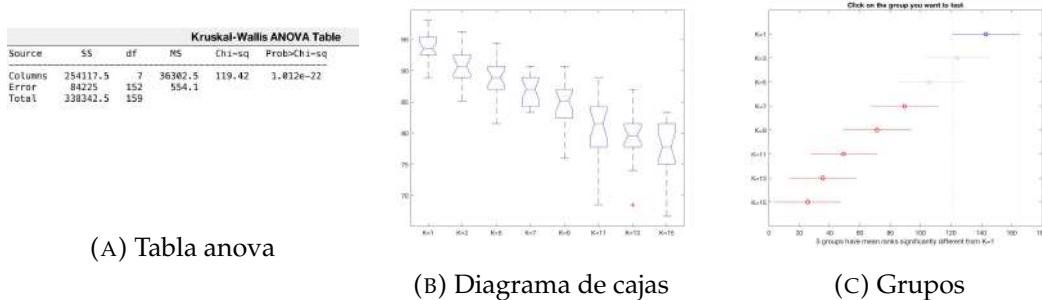


FIGURA D.9: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 2 dimensiones del espacio de color CIE L*a*b*

D.2.4. Utilizando histogramas de 3 dimensiones del espacio de color CIE L*a*b*

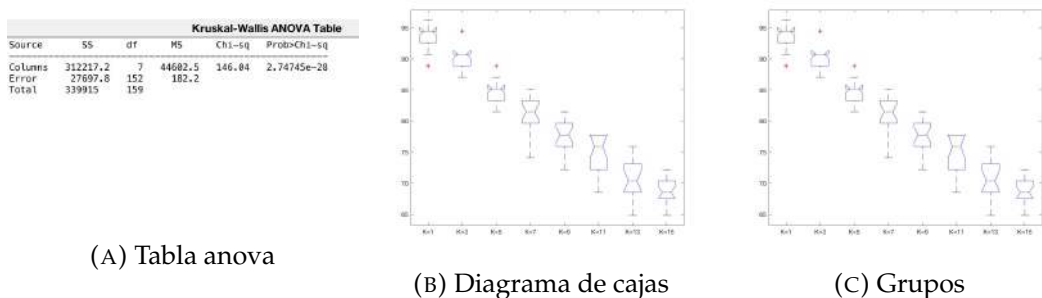


FIGURA D.10: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 3 dimensiones del espacio de color CIE L*a*b*

D.2.5. Utilizando la unión de los histogramas de 2 dimensiones de los espacios de color HSI con CIE L*a*b*

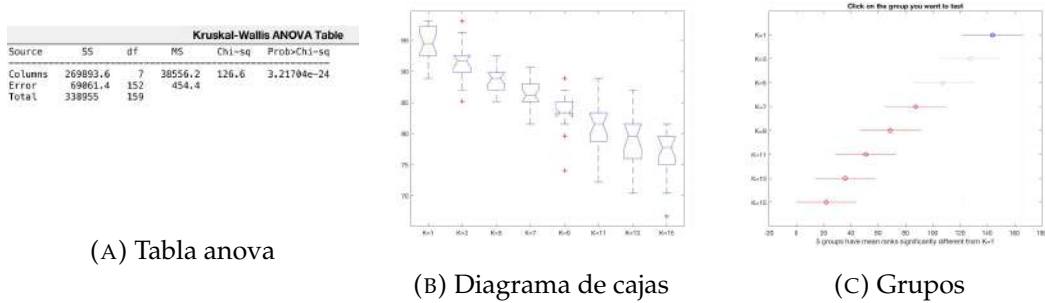


FIGURA D.11: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 2 dimensiones de los espacios de color HSI con CIE L*a*b*

D.2.6. Utilizando la unión de los histogramas de 3 dimensiones de los espacios de color HSI con CIE L*a*b*

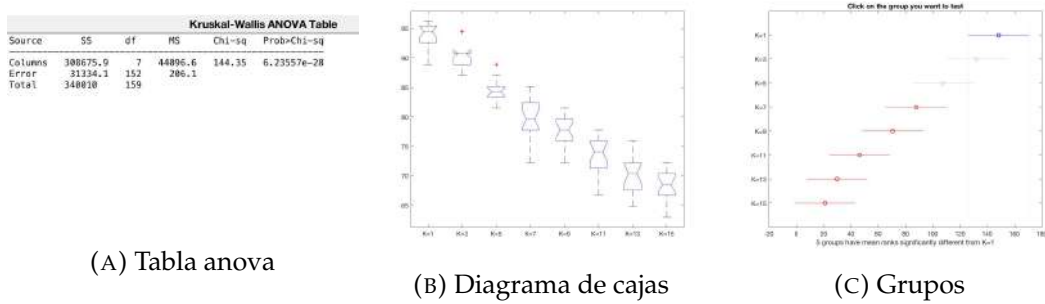


FIGURA D.12: Gráficas resultantes de pruebas estadísticas de los resultados de la clasificación de poblaciones nativas de frijol por color del grano utilizando histogramas de 3 dimensiones de los espacios de color HSI con CIE L*a*b*

Referencias

- [1] 2017. URL: <https://www.scratchapixel.com/lessons/digital-imaging/colors/color-space>.
- [2] 2017. URL: https://www.probabilitycourse.com/chapter5/5_1_1_joint_pmf.php.
- [3] Elia Nora Aquino-Bolaños, Yatzil Denih García-Díaz, José Luis Chavez-Servia, Jose Cruz Carrillo-Rodríguez, Araceli Minerva Vera-Guzmán, Elena Heredia-García. «Anthocyanin, polyphenol, and avonoid contents and antioxidant activity in Mexican common bean (*Phaseolus vulgaris* L.) landraces». En: *Emirates Journal of Food and Agriculture* 28 (2016), págs. 581-588.
- [4] Ethem Alpaydın. «Introduction to Machine Learning». En: The MIT Press Cambridge, Massachusetts London, England,[Google book version], 2014. URL: <https://books.google.com.mx/books?id=NP5bBAAAQBAJ&lpg=PP1&hl=es&pg=PR7#v=onepage&q&f=false>.
- [5] Sidnei Alves de AraÃjo, Jorge Henrique Pessota y Hae Yong Kim. «Beans quality inspection using correlation-based granulometry». En: *Engineering Applications of Artificial Intelligence* 40 (2015), págs. 84 -94. ISSN: 0952-1976. DOI: <http://dx.doi.org/10.1016/j.engappai.2015.01.004>. URL: <http://www.sciencedirect.com/science/article/pii/S0952197615000056>.
- [6] Asrat Asfaw, Matthew W. Blair y Conny Almekinders. «Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands». En: *Theoretical and Applied Genetics* 120.1 (2009), págs. 1-12. ISSN: 1432-2242. DOI: [10.1007/s00122-009-1154-7](https://doi.org/10.1007/s00122-009-1154-7). URL: <https://doi.org/10.1007/s00122-009-1154-7>.
- [7] Baohua Zhang and Wenqian Huang and Jiangbo Li and Chunjiang Zhao and Shuxiang Fan and Jitao Wu and Chengliang Liu. «Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review». En: *Food Research International* 62.Supplement C (2014), págs. 326 -343. ISSN: 0963-9969. DOI: <https://doi.org/10.1016/j.foodres.2014.03.012>. URL: <http://www.sciencedirect.com/science/article/pii/S0963996914001707>.
- [8] Bitjoka Laurent and Boukar Ousman and Tenin Dzudie and MF Mbofung Carl and Tonye Emmanuel. «Digital camera images processing of hard-to-cook beans». En: *Journal of Engineering and Technology Research* 2(9).9 (2010), págs. 177-188.
- [9] Elena Bitocchi y col. «Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data». En: *Proceedings of the National Academy of Sciences* 109.14 (2012), E788-E796. DOI: [10.1073/pnas.1108973109](https://doi.org/10.1073/pnas.1108973109). eprint: <http://www.pnas.org/content/109/14/E788.full.pdf>. URL: <http://www.pnas.org/content/109/14/E788.abstract>.

- [10] Celia Chávez-Mendoza y Esteban Sánchez. «Bioactive Compounds from Mexican Varieties of the Common Bean (*Phaseolus vulgaris*): Implications for Health». En: *Molecules* 22.8 (2017). ISSN: 1420-3049. DOI: [10.3390/molecules22081360](https://doi.org/10.3390/molecules22081360). URL: <http://www.mdpi.com/1420-3049/22/8/1360>.
- [11] Chávez-Servia, José Luis and Heredia-García, Elena and Mayek-Pérez, Netzahualcoyotl and Aquino-Bolaños, Elia N. and Delgado, Sanjuana Hernández- and Carrillo-Rodríguez, José C. and Gill-Langarica, Homar R. and Vera-Guzmán, Araceli M. «Diversity of Common Bean (*Phaseolus vulgaris* L.) Landraces and the Nutritional Value of their Grains». En: *Grain Legumes*. Ed. por Aakash Kumar Goyal. Rijeka: InTech, 2016. Cap. 01. DOI: [10.5772/63439](https://doi.org/10.5772/63439). URL: <http://dx.doi.org/10.5772/63439>.
- [12] P. Gepts y D Debouck. *Origin, domestication, and evolution of the common bean (Phaseolus vulgaris L.)* Ed. por Aart van Schoonhoven y O. Voysest. Vol. Common Beans: Research for Crop Improvement. 7-53. Cali, Colombia, 1991.
- [13] Jian Gong, Liyuan Li y Weinan Chen. «Fast recursive algorithms for two-dimensional thresholding». En: *Pattern Recognition* 31.3 (1998), págs. 295 -300. ISSN: 0031-3203. DOI: [https://doi.org/10.1016/S0031-3203\(97\)00043-5](https://doi.org/10.1016/S0031-3203(97)00043-5). URL: <http://www.sciencedirect.com/science/article/pii/S0031320397000435>.
- [14] R.C. Gonzalez y R.E. Woods. *Digital image processing*. Prentice Hall, 2002.
- [15] P.H. Graham y p. Ranalli. «Common bean (*Phaseolus vulgaris* L.)» En: *Field Crops Research, elsevier* 53 (1997), págs. 131-146.
- [16] Salvador Guzmán-Maldonado y Octavio Paredes-López. «Perdidas irreversibles de un rico legado de la naturaleza para la humanidad. El caso del frijol-comun en Mexico». En: (dic. de 1998).
- [17] Copyright ©2000 Adobe Systems Incorporated. *Color Models CIELAB*. Ed. por Technical guides. 2017. URL: http://dba.med.sc.edu/price/irf/Adobe_tg/models/cielab.html.
- [18] J. Li and G. Liao and Z. Ou and J. Jin. «Rapeseed Seeds Classification by Machine Vision». En: *Workshop on Intelligent Information Technology Application (IITA 2007)*. 2007, págs. 222-226. DOI: [10.1109/IITA.2007.56](https://doi.org/10.1109/IITA.2007.56).
- [19] K. Kiratiratanapruk and W. Sinthupinyo. «Color and texture for corn seed classification by machine vision». En: *2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*. 2011, págs. 1-5. DOI: [10.1109/ISPACS.2011.6146100](https://doi.org/10.1109/ISPACS.2011.6146100).
- [20] Inc. (2003) Konica Minolta Holdings. *Konica Minolta Holdings*. 26 de julio 2017. 2017. URL: <http://www2.konicaminolta.eu/eu/Measuring/pcc/es/part3/01.html>.
- [21] Laurent Bitjoka and Ousman Boukar and Alban Ngatchou and Gamraïkréo Djaowé and Lamsikreo Banbe. «Fast Objective Identification of Beans Grains (*Phaseolus vulgaris* L.) Varieties Using CFA Images Compacity Measurement». En: *Electrical and Electronic Engineering* 5 (2015), págs. 1-5.
- [22] M. You and C. Cai. «Weed Seeds Classification Based on PCA, 2DPCA, Column-directional 2DPCA and (2D)2PCA». En: *2009 International Asia Symposium on Intelligent Interaction and Affective Computing*. 2009, págs. 187-190. DOI: [10.1109/ASIA.2009.57](https://doi.org/10.1109/ASIA.2009.57).

- [23] Marisol Lo Bianco and Oscar Grillo and Roberto Cremonini and Marco Sarigu and Gianfranco Venora. «Characterisation of Italian bean landraces (*Phaseolus vulgaris* L.) using seed image analysis and texture descriptors». En: *Australian Journal of Crop Science* 9 (2015), págs. 1022-1034.
- [24] Raymond G. McGuire. *Reporting of Objective Color Measurements*. Inf. téc. U.S. Department of Agriculture-Agricultural Research Service, 1992. URL: <http://hortsci.ashspublications.org/content/27/12/1254.full.pdf>.
- [25] A. Nasirahmadi y N. Behrooz-Khazaei. «Identification of bean varieties according to color features using artificial neural network». En: *Spanish Journal of Agricultural Research* 11.3 (2013), págs. 670-677.
- [26] Seyed Saeid Modarres Najafabadi and Leila Farahani. «Shape analysis of common bean (*Phaseolus vulgaris* L.) seeds using image analysis». En: *International Research Journal of Applied and Basic Sciences* Vol., 3 (8).ISSN 2251-838X (2012), págs. 1619-1623.
- [27] Diego Furtado Silva y Laboratório De Inteligência Computacional (labic. «How k-Nearest Neighbor Parameters Affect its Performance». En: *Simposio Argentino de Inteligencia Artificial (ASAI 2009)*, 95 – 106. 2009.
- [28] Jun Tang. «A color image segmentation algorithm based on region growing». En: *2010 2nd International Conference on Computer Engineering and Technology*. Vol. 6. 2010, págs. V6-634-V6-637. DOI: [10.1109/ICCET.2010.5486012](https://doi.org/10.1109/ICCET.2010.5486012).
- [29] Apolonio Vargas-Torres y col. «Chemical Composition, Starch Bioavailability and Indigestible fraction of Common Beans (*Phaseolus Vulgaris* L.)» En: *Starch - Stärke* 56.2 (2004), págs. 74-78. ISSN: 1521-379X. DOI: [10.1002/star.200300205](https://doi.org/10.1002/star.200300205). URL: <http://dx.doi.org/10.1002/star.200300205>.
- [30] G. Venora y col. «Identification of Sicilian landraces and Canadian cultivars of lentil using an image analysis system». En: *Food Research International* 40.1 (2007), págs. 161 -166. ISSN: 0963-9969. DOI: <https://doi.org/10.1016/j.foodres.2006.09.001>. URL: <http://www.sciencedirect.com/science/article/pii/S0963996906001499>.
- [31] Gianfranco Venora y col. «Identification of Italian landraces of bean (*Phaseolus vulgaris* L.) using an image analysis system». En: *Scientia Horticulturae* 121.4 (2009), págs. 410 -418. ISSN: 0304-4238. DOI: <http://dx.doi.org/10.1016/j.scienta.2009.03.014>. URL: <http://www.sciencedirect.com/science/article/pii/S0304423809001423>.
- [32] Yan, Xiaomei and Wang, Jinxing and Liu, Shuangxi and Zhang, Chunqing", editor="Li, Daoliang and Liu, Yande and Chen, Yingyi. «Purity Identification of Maize Seed Based on Color Characteristics». En: *Computer and Computing Technologies in Agriculture IV: 4th IFIP TC 12 Conference, CCTA 2010, Nanchang, China, October 22-25, 2010, Selected Papers, Part III*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, págs. 620-628. ISBN: 978-3-642-18354-6. DOI: [10.1007/978-3-642-18354-6_73](https://doi.org/10.1007/978-3-642-18354-6_73). URL: https://doi.org/10.1007/978-3-642-18354-6_73.