



UNIVERSIDAD NACIONAL DE COLOMBIA

Clasificación de las estructuras vegetativas presentes en ramas de café usando visión de máquina

Jonathan Avendaño Pérez

Universidad Nacional de Colombia Facultad de Ingeniería

Departamento de Ingeniería Electrónica

Bogotá, Colombia

2017

Clasificación de las estructuras vegetativas presentes en ramas de café usando visión de máquina

Jonathan Avendaño Pérez

Tesis de grado presentada como requisito parcial para optar al título de:
Magister en Ingeniería Electrónica

Director:

Flavio Augusto Prieto Ortiz, Ph.D.

Co-directora:

Paula Jimena Ramos Giraldo, M.Sc.

Línea de Investigación: Visión Artificial

Grupo de Investigación: GAUNAL

Universidad Nacional de Colombia
Facultad de Ingeniería
Departamento de Ingeniería Electrónica
Bogotá, Colombia
2017

Dedico el siguiente trabajo a mi hijo Sebastián,
a mis padres y hermano; y a mi compañera de
vida Mónica.

Agradecimientos

Agradezco en primera instancia Dios por la sabiduría que me brindó y por permitirme realizar este trabajo. A todas las personas que de alguna manera me apoyaron para alcanzar este objetivo les agradezco con todo mi corazón. Muchas gracias a mi director, el profesor Flavio Prieto, por toda la ayuda y enseñanzas que me brindó durante mis estudios de maestría en la Universidad Nacional de Colombia. A mi co-directora, la magíster Paula Ramos, le agradezco por su dedicación y apoyo al desarrollo de este trabajo.

También, quiero agradecer enormemente al Centro Nacional de Investigaciones de Café (Cenicafé), quienes proporcionaron personal para la adquisición de las imágenes en campo al igual que todos los datos necesarios para la realización del proyecto. De igual forma quiero agradecer a Colciencias por el financiamiento recibido durante el desarrollo del proyecto.

Finalmente Agradezco a la Universidad Nacional de Colombia por permitirme iniciar y tomar mis estudios de maestría en sus instalaciones.

A todos ellos muchas gracias.

Resumen

El café es una de las principales bebidas a nivel mundial y como producto agrícola requiere herramientas para monitorear y controlar el cultivo de manera no destructiva. En búsqueda de desarrollar sistemas no destructivos que trabajen en campo, en esta tesis se presenta el desarrollo de un sistema para identificar y clasificar seis estructuras vegetativas (hojas, tallos, flores, frutos inmaduros, semi-maduros, y maduros) presentes en las ramas de café. Se adquirieron vídeos con la cámara posterior de un dispositivo móvil, de 12 ramas de café en condiciones de campo. Por cada vídeo se seleccionaron los frames con mayor información acerca de la escena. Luego por cada rama se generó una reconstrucción 3D utilizando las técnicas de *Structure Form Motion* (SFM) y *Patch-based Multi-view Stereo* (PMVS). Todas las imágenes adquiridas fueron anotadas manualmente y posteriormente una nube de puntos (*GroundTruth*) fue generada por cada rama. Las nubes de puntos generadas fueron filtradas utilizando un filtrado estadístico de *outliers*, con el objetivo de eliminar el ruido generado en el proceso de reconstrucción 3D. Los puntos que se encontraban a mayor profundidad se consideraron el fondo de la escena y fueron removidos utilizando un filtro passa-banda. Las nubes de puntos fueron sub-muestreadas mediante un proceso de voxelización, con el objetivo de reducir el número de puntos en un 50% y por ende reducir el costo computacional de los siguientes procesos. Características 2D y 3D fueron extraídas de las nubes de puntos, 11 características basadas en los espacios de color RGB, Lab, Luv, YCbCr y HSV, 4 características basadas en curvaturas, y los índices de forma y curvatura. Una máquina de vectores de soporte (SVM) fue entrenada con las características previamente encontradas, utilizando 8 ramas para la fase de entrenamiento y 4 para la fase de validación. Los resultados experimentales mostraron una precisión del 0.82 y un recall 0.79, clasificando las diferentes estructuras vegetativas. El sistema propuesto es económico ya que solo se necesita de un dispositivo móvil para la captura de la información; sin embargo, los demás procesos del sistema se realizan de manera *offline*. Adicionalmente, el sistema desarrollado mostró ser robusto a los cambios de luminosidad presentes durante la captura de los vídeos en campo.

Abstract

As a drink, coffee is one of the most in demand products worldwide; as an agricultural product, it requires non-destructive tools for monitoring and control. With the purpose of creating a non-destructive system which can be used in the field, a system was developed to find and classify six types of vegetative structures on coffee branches: leaves, stems, flowers, unripe fruits, semi-ripe fruits, and ripe fruits. Videos were obtained from 12 coffee branches in field conditions, using the rear camera of a mobile device. Approximately 90 frames, those which had the largest amount of information from the scene, were selected from each video; then, a three-dimensional (3D) reconstruction was generated using the Structure from Motion (SfM) and Patch-based Multi-view Stereo (PMVS) techniques per each branch. All of the acquired images were manually recorded and then, a Ground Truth point cloud was generated for each branch. The generated point clouds were filtered using a statistical outliers filter, in order to eliminate noise generated in the 3D reconstruction process. The points that were located in the deepest part were taken as the background of the scene, and were removed using a band-pass filter. The point clouds were sub-sampled using a VoxelGrid filter, to reduce the number of points to 50% and, therefore, reduce the computation time of the following processes. Some two-dimensional (2D) and 3D features were taken from the point clouds: 11 of them based on RGB, Lab, Luv, YCbCr, and HSV color space, 4 of them based on curvatures, and the remaining two based on shape and curvedness indexes. A Support Vector Machine (SVM) was trained with the previously encountered features by using eight branches for the training stage, and four branches for the validation stage. The experimental results showed a precision of 0.82 and a recall of 0.79 when classifying the different vegetative structures. The proposed system is economical since only a mobile device is needed to obtain the information; nonetheless, the remaining processes of the system were performed offline. Additionally, the developed system was not affected by changes in lighting conditions, when recording the videos on a coffee plantation.

Contenido

Agradecimientos	vii
1. Introducción	2
1.1. Trabajos Relacionados	3
1.2. Contribución	5
2. Fundamentación Teórica	7
2.1. Generalidades del Café	7
2.1.1. El arbusto	7
2.1.2. Estructuras vegetativas presentes en ramas de café	8
2.2. Modelo de Cámara	12
2.2.1. Geometría proyectiva	12
2.2.2. Distorsión de lentes	16
2.3. Calibración de cámara	18
2.4. Estructura a partir de movimiento	21
2.4.1. Extracción de puntos característicos utilizando SIFT	22
2.4.2. Emparejamiento de puntos característicos utilizando Flujo Óptico	24
2.4.3. Cálculo de la matriz fundamental	26
2.4.4. Cálculo del movimiento de la cámara	26
2.4.5. Triangulación	27
2.5. Máquinas de vectores de soporte	28
3. Clasificación de las estructuras vegetativas en ramas de café	31
3.1. Datos de entrada	32
3.1.1. Adquisición de las imágenes	32
3.1.2. Anotación manual de las imágenes capturadas en campo	33
3.1.3. Generación de la Nube de puntos Ground Truth	33
3.2. Reconstrucción 3D	35
3.3. Filtrado	39
3.3.1. Filtrado estadístico de <i>outliers</i>	39
3.3.2. Remoción de Fondo	41
3.4. Sub-muestreo	42
3.5. Extracción de características	46
3.5.1. Características 3D	46

3.5.2. Características 2D	47
3.6. Clasificación	48
4. Resultados	50
4.1. Índices de evaluación	52
4.2. Clasificación de las estructuras vegetativas	52
4.3. Discusión	63
5. Conclusiones y Trabajos Futuros.	66
Bibliografía	68

Lista de Figuras

2-1. Arbusto de café Variedad Castillo®. Imagen tomada de la página web de Cenicafé.	7
2-2. Rama de <i>Coffea arabica</i> . Imagen tomada de la base de dibujos de Cenicafé. .	8
2-3. Etapas de desarrollo de los frutos de café. [1]	9
2-4. Tallo y ramas principales de un cafeto. Imagen tomada de la base de imágenes de Café de Colombia.	10
2-5. Hojas de un cafeto. Imagen tomada de la base de imágenes de Café de Colombia.	10
2-6. Flor de cafeto. Imagen tomada de la base de imágenes de Café de Colombia.	10
2-7. Estado de desarrollos: inmaduro, semi-maduro y maduro.	11
2-8. Geometría cámara <i>pinhole</i> . Imagen tomada de [2].	12
2-9. Sistema de coordenadas en la imagen (x, y) y en la cámara (x_{cam}, y_{cam}) . Imagen tomada de [2].	14
2-10. Transformación euclídea entre las coordenadas del mundo real y las coordenadas de la imagen. Imagen tomada de [2].	15
2-11. Distorsión radial. Imagen tomada de [3].	16
2-12. Distorsión tangencial. Imagen tomada de [3].	17
2-13. Proceso de adquisición imágenes desde diferentes orientaciones del patron planar para el proceso de calibración de cámara. Imagen tomada de [3].	18
2-14. Error geométrico presente en el proceso de calibración de cámara.	20
2-15. Estructura a partir de movimiento. Tomado de [4]	21
2-16. Diferencia de Gaussiano en diferentes escalas de la imagen. Tomado de [5] . .	23
2-17. Extracción de <i>keypoints</i> utilizando SIFT y emparejamiento por medio de flujo óptico	25
2-18. Triangulación de puntos. Tomado de [2]	27
2-19. Ejemplo de varios hiper-planos de separación con sus respectivos márgenes para una SVM.	28
2-20. Ejemplo arquitectura de una SVM. Adaptada de [6]	30
3-1. Sistema propuesto para clasificación de estructuras vegetativas presentes en ramas de café.	32
3-2. Ejemplo de adquisición de imágenes en un cafeto.	33
3-3. Imágenes adquiridas junto con su respectiva anotación.	34

3-4. Ejemplo nubes de puntos GroundTruth generadas.	35
3-5. Patrón utilizado para la fase de calibración de cámara.	35
3-6. Proceso usado para la implementación de SfM.	36
3-7. <i>Keypoints</i> encontrados utilizando el algoritmo de SIFT.	37
3-8. Emparejamiento de los <i>keypoints</i>	37
3-9. Nube de puntos generada por SfM.	38
3-10. Definición de un parche (patch) y de las imágenes asociadas con el parche. Tomado de [7].	38
3-11. Nube de puntos generada por PMVS	39
3-12. Filtrado Estadístico de <i>Outliers</i> sobre un vecindario.	40
3-13. Fase de Filtrado.	41
3-14. Sub-muestreo utilizando filtrado por Voxelización	42
3-15. Fase de sub-muestreo.	43
3-16. Resultados del proceso de filtrado de la rama No. 1	43
3-17. Resultados del proceso de filtrado de la rama No. 2	43
3-18. Resultados del proceso de filtrado de la rama No. 3	44
3-19. Resultados del proceso de filtrado de la rama No. 4	44
3-20. Resultados del proceso de sub-muestreo en la rama No. 1.	44
3-21. Resultados del proceso de sub-muestreo en la rama No. 2.	45
3-22. Resultados del proceso de sub-muestreo en la rama No. 3.	45
3-23. Resultados del proceso de sub-muestreo en la rama No. 4.	45
3-24. Procesos para la extracción de regiones y cálculo de curvaturas.	46
3-25. Extracción de Características 2D.	48
3-26. Proceso de validación cruzada usando un tamaño de región = 2.5cm.	49
4-1. Ramas utilizadas para la validación del sistema.	51
4-2. Matriz de confusión resultante al utilizar un tamaño de región de 0.5cm	53
4-3. Matriz de confusión resultante al utilizar un tamaño de región de 1.0cm	54
4-4. Matriz de confusión resultante al utilizar un tamaño de región de 1.5cm	54
4-5. Matriz de confusión resultante al utilizar un tamaño de región de 2.0cm	55
4-6. Matriz de confusión resultante al utilizar un tamaño de región de 2.5cm	55
4-7. Matriz de confusión resultante al utilizar un tamaño de región de 3.0cm	56
4-8. Matriz de confusión resultante al utilizar un tamaño de región de 4.0cm	56
4-9. Matriz de confusión resultante al utilizar un tamaño de región de 5.0cm	57
4-10. Análisis del resultado de anotación fruto semi-maduro.	59
4-11. Tamaño región vs. Índices de evaluación.	60
4-12. Nubes de puntos GroundTruth vs. los resultados obtenidos por el sistema propuesto.	61
4-13. Matriz de confusión uniendo las clases: fruto inmaduro, semi-maduro y maduro. . . .	62

Lista de Tablas

3-1. Parámetros intrínsecos de la cámara.	36
3-2. Tamaño de las nubes de puntos en cada fase del proceso de filtrado y sub- muestreo	46
4-1. Porcentaje de datos utilizado por clase en las fases de entrenamiento y validación.	50
4-2. F1 - Score por clase vs. tamaño de región	53
4-3. Tiempo empleado por rama en cada etapa del sistema propuesto.	62
4-4. Tabla comparativa de algunos trabajos relacionados con el método propuesto.	65

1 Introducción

El café es una de las principales bebidas a nivel mundial, se obtiene a partir de los frutos procesados de la planta de café. Para obtener una tasa de café son necesarias varias fases dentro de las que se encuentran: la siembra, el cuidado del cultivo, la cosecha y la poscosecha. Colombia es uno de los principales exportadores de café a nivel mundial. El café colombiano se caracteriza por su suavidad y textura. Adicionalmente, el café de Colombia es reconocido por su proceso de recolección, el cual se realiza de manera manual y busca seleccionar únicamente los granos maduros durante la recolección para obtener la mejor calidad. En Colombia, diferentes zonas del territorio nacional, como por ejemplo el eje cafetero, producen café debido a que es un país con clima tropical. Las condiciones ambientales ideales para un cultivo de café según la Federación Nacional de Cafeteros de Colombia se encuentran entre los 1200 y 1800 metros de altura sobre el nivel del mar, con temperaturas que oscilan entre los 17 y los 23 grados Celsius y con precipitaciones cercanas a los 2000 milímetros anuales. Las condiciones climáticas en las que se cultiva el café influyen en la calidad del mismo. Sin embargo, la calidad del café colombiano también depende de la variedad de café utilizada, las prácticas de cultivo implementadas, el cuidado de la plantación, y los procesos posteriores a la recolección.

Actualmente, el Centro de Investigación Nacional de Café (Cenicafé), entrega a los caficultores de Colombia técnicas y tecnologías para tener cultivos sanos y productivos; adicionalmente, tener procesos de cosecha y poscosecha más eficientes y amigables ambientalmente. Lo anterior permite a los productores de café en Colombia ser competitivos a nivel mundial, y tener una oferta de café de buena calidad.

En la actualidad, en Colombia, se han desarrollado diversas investigaciones que usan técnicas de visión de máquina y se aplican al café, pero muchas de estas aún no están implementadas [8, 9]. En Colombia tan sólo se han implementado sistemas de visión artificial que detectan y separan granos de café en almendra [10]. Existe la necesidad de desarrollar sistemas no destructivos que estimen el estado desarrollo de los granos de café y, posteriormente, permitan realizar pronósticos y proyecciones de cosecha a nivel de finca que ayuden a los caficultores a tomar decisiones.

El trabajo presentado en este documento, está enmarcado dentro del proyecto cofinanciado por Colciencias titulado: Desarrollo y validación de una herramienta tecnológica basada en

visión de máquina y TICs para la administración de labores agrícolas en la finca cafetera, donde la Universidad Nacional de Colombia y Cenicafé han unido esfuerzos para desarrollar un sistema de visión de máquina que estime la producción de café en finca, y genere un impacto positivo en la caficultura colombiana, al ayudar a los caficultores en la toma de decisiones para tiempos óptimos de cosecha.

Este trabajo desarrolló un sistema de clasificación de estructuras vegetativas presentes en ramas de café, a partir de características 2D y 3D extraídas de videos tomados en campo bajo condiciones de luminosidad no controladas. La herramienta desarrollada mostró independencia a las condiciones de luminosidad, ruido y oclusión presentes cuando se trabajan con imágenes en campo, y es un punto de partida para la generación de nuevas tecnologías que apoyen las labores agrícolas en Colombia, especialmente a los caficultores colombianos.

1.1. Trabajos Relacionados

El desarrollo de sistemas expertos aplicados en la agricultura, es hoy en día uno de los temas de interés para las diferentes áreas de las ciencias de la computación. Los sistemas de visión de máquina aplicados en la agricultura han permitido generar tecnologías no destructivas que realizan tareas agrícolas en campo, granjas, invernaderos, algunos ejemplos de estos trabajos son: la clasificación de suelos para seleccionar las técnicas de labrado a utilizar en cultivos [11]; el desarrollo de robots agrícolas [12, 13], la inspección de calidad y detección de enfermedades en alimentos [14, 15, 16, 17, 18]; el reconocimiento del tipo de fruto utilizado y la determinación del tiempo de envejecimiento en licores [19]; la medición del volumen en productos alimenticios como es el caso de las manzanas deshidratadas [20] y el monitoreo de los granos de café en la fase de tostado [21].

Las principales técnicas utilizadas en las tareas listadas anteriormente se relacionan con visión de máquina y buscan hacer más eficientes los procesos productivos agropecuarios. En la actualidad existen desafíos vigentes para el procesamiento digital de imágenes, por ejemplo: detectar y realizar un seguimiento de rupturas estructurales en organismos y células en los cultivos, modelar automáticamente la geometría 3D de las plantas, segmentar y vigilar bajo tierra las raíces de las plantas, seguir el crecimiento de los cultivos y realizar una predicción acertada de cosechas [22].

Diferentes investigaciones utilizando el procesamiento de imágenes han desarrollado sistemas de visión artificial para la clasificación de estructuras vegetativas, por ejemplo los desarrollados para: hortalizas [23], uvas [24], dátiles [25], tomate [26], manzanas [20, 16] e incluso café [27, 28, 29, 30]. Los trabajos anteriores se desarrollaron bajo ambientes de luminosidad controlados. Sin embargo; uno de los principales inconvenientes en el desarrollo de los sistemas de visión artificial aplicados al agro, es construir soluciones que sean independientes a

las variaciones de luminosidad en las que se presenta una escena. Las sombras, el ruido, y los reflejos o brillos excesivos son problemas que se derivan de la iluminación [31]. Por ejemplo, al trabajar con imágenes capturadas en campo, no es posible controlar las condiciones de luminosidad que varían según la hora del día en la cual se adquieren los datos. Lo anterior genera inconvenientes cuando se procesa información de color y textura (2D), debido a que se presentan valores atípicos en las características extraídas de las imágenes, las cuales no permiten generar patrones para la identificación de las estructuras presentes en la escena. Adicionalmente, a los problemas ocasionados por la iluminación al trabajar con imágenes capturadas en campo se suman los inconvenientes relacionados con la oclusión de las estructuras vegetativas presentes en el cultivo. Las oclusiones se presentan debido a la naturaleza de los cultivos y la forma en que son adquiridas las imágenes [32, 33]. La implementación de técnicas basadas en características de forma sobre imágenes de campo genera cierta independencia a los problemas de oclusión [34, 35].

Existen otras tecnologías para adquirir información 3D en campo, por ejemplo: visión estéreo [36] y láser [32]. Las anteriores soluciones aunque permiten el uso de información 3D, en ocasiones resultan ser costosas económicamente. Una manera económica de capturar información 3D es mediante el uso de la técnica de SfM, donde se pueden reconstruir estructuras vegetativas en 3D de diferentes cultivos, debido a que se necesita únicamente de una cámara para la captura de información. Sin embargo, sí las estructuras vegetativas del cultivo de estudio son muy livianas, se pueden presentar inconvenientes con el movimiento de las estructuras por el viento [37], pero este no es el caso de los cultivos de café.

Recientemente, técnicas no destructivas usando visión por computador han sido implementadas para la detección y clasificación de estructuras vegetativas, trabajando con imágenes de campo. Los diferentes trabajos han sido aplicados en gran variedad de cultivos como: maíz [38, 39, 40], cereal [41], tomate [35] y naranjas [42]. Los anteriores trabajos se basan en el uso de características 2D. Por su parte trabajos que utilizan técnicas 3D han sido implementados en cultivos de: duraznos [43], uvas [44], algodón [45], piñas [46], y hortalizas como col, repollo, coliflor, girasol y remolacha [37]. Finalmente, algunos trabajos utilizan conjuntamente sensores 2D y 3D como es el caso del trabajo presentado en [36], donde se realizó un sistema para el conteo automático de manzanas. Por lo anterior se evidencia una tendencia a investigar y desarrollar sistemas no destructivos que trabajen en campo para la clasificación de estructuras vegetativas.

A diferencia de los trabajos relacionados anteriormente, este trabajo consistió en el desarrollo de un sistema para la clasificación de estructuras vegetativas presentes en ramas de café, que trabaja con imágenes en movimiento capturadas en campo por un sistema monocular (cámara principal de un dispositivo móvil), bajo las condiciones de luminosidad ofrecidas por el ambiente en el momento que se capturan las imágenes. Adicionalmente, se trabajó con

estructuras cuyo tamaño oscila de 1 a 4 centímetros y se clasificaron 6 clases de estructuras vegetativas, tallo, hojas, flores y frutos en tres diferentes estados de maduración: inmaduro, semi-maduro y maduro. Se propone el uso de características 3D para mitigar el efecto negativo que tienen los cambios de iluminación y oclusión, sobre los sistemas de clasificación que se basan únicamente en características 2D. La información de color desempeña un papel importante, a la hora de clasificar las estructuras vegetativas de las ramas de café. Por ejemplo, la forma y geometría de los granos de café en sus diferentes estados de maduración pueden llegar a ser similares, por lo que es necesario integrar características 2D que permitan la separación entre las clases. El trabajo propuesto utiliza características 2D y 3D, por lo tanto no es necesario la implementación de varios sensores, lo que hace que la solución propuesta sea económica, debido a que para la adquisición de los datos se necesita únicamente un dispositivo móvil, como por ejemplo la cámara principal de un teléfono inteligente. El sistema propuesto logró tener independencia a las condiciones de luminosidad y oclusión. Aunque se obtuvo una precisión cercana al 82 %, existen clases que debido a su similitud en forma y color presentan inconvenientes para su correcta clasificación.

1.2. Contribución

La contribución de esta tesis fue el desarrollo un sistema para clasificar seis estructuras vegetativas (tallo, hoja, flor, fruto inmaduro, fruto semi-maduro y fruto maduro), presentes en ramas de café a partir de vídeos tomados bajo condiciones de campo. El sistema desarrollado forma parte de una herramienta informática que pretende romper la brecha que existe entre la tecnología y los caficultores de Colombia. El sistema que se presenta en este trabajo abordó diferentes desafíos y se resume a continuación:

- El sistema desarrollado mostró independencia a cambios de luminosidad los cuales se presentan cuando se trabaja con imágenes de campo. Los problemas por cambios de iluminación fueron resueltos mediante la generación estructuras 3D de las ramas de café a partir de múltiples vistas, esto ayudó a que la información de color, la cual se ve afectada por sombras y excesos de luminosidad, no dependa de un único dato sino de varios. Adicionalmente, se utilizaron características 3D basadas en curvaturas, las cuales no dependen de la información de color para la clasificación de las estructuras.
- Se mitigó los problemas de oclusión de las estructuras vegetativas presentes en cultivos de café. Al utilizar información 3D fue posible trabajar con la ubicación espacial de los frutos, lo que permitió identificar frutos aun cuando estos estaban ocluidos por otras estructuras.
- Se logró realizar una separación del fondo de las imágenes el cual estaba conformado por suelo, malas hierbas, hojas secas, y ramas de otros árboles. Lo anterior fue solucionado

utilizando filtros pasa banda sobre las estructuras 3D generadas, esto permitió trabajar únicamente con la información de interés en la escena y mitigar el error causado por el fondo.

- Se obtuvo una herramienta no destructiva para la clasificación de las estructuras vegetativas presentes en ramas de café. El sistema presentado utiliza únicamente para la captura de información la cámara principal de un teléfono inteligente. Lo anterior permite que la solución propuesta en este trabajo sea una herramienta accesible para los caficultores de Colombia.

2 Fundamentación Teórica

2.1. Generalidades del Café

El cafeto es el nombre que recibe el arbusto del café. La especie del café producido en Colombia corresponde a *Coffea arabica*. Borbón, Maragogipe, Tabi, Típica, y las variedades de porte bajo: caturra, variedad Colombia y variedad Castillo®, son algunas de las variedades que se cultivan en Colombia.

2.1.1. El arbusto

En este trabajo se utilizaron datos obtenidos en cultivos Variedad Castillo®. Esta variedad de café fue liberada por Cenicafé en el año 2005, y fue obtenida por medio de estrategias de mejoramiento genético, con el objetivo de mejorar la producción, dar resistencia a las roya, probable tolerancia a la enfermedad de las cerezas, excelentes atributos agronómicos, mayor tamaño del grano y calidad en taza [47].

La Figura 2-1 ilustra un arbusto de café Variedad Castillo®, los cuales alcanzan alturas de 2 metros y florecen a partir del mes 12 después de la siembra. El tallo de arbusto mide entre 7 y 8 centímetros de diámetro y el tiempo de producción para esta variedad de café es de 30 a 36 semanas dependiendo de las condiciones ambientales del cultivo.



Figura 2-1: Arbusto de café Variedad Castillo®. Imagen tomada de la página web de Cenicafé.

2.1.2. Estructuras vegetativas presentes en ramas de café

La Figura 2-2, ilustra las diferentes estructuras vegetativas que se presentan en las ramas de *Coffea arabica*. Se puede observar que existen 8 estructuras: rama, nudo, flor, hoja, pedúnculo, pergamino, almendra y fruto.

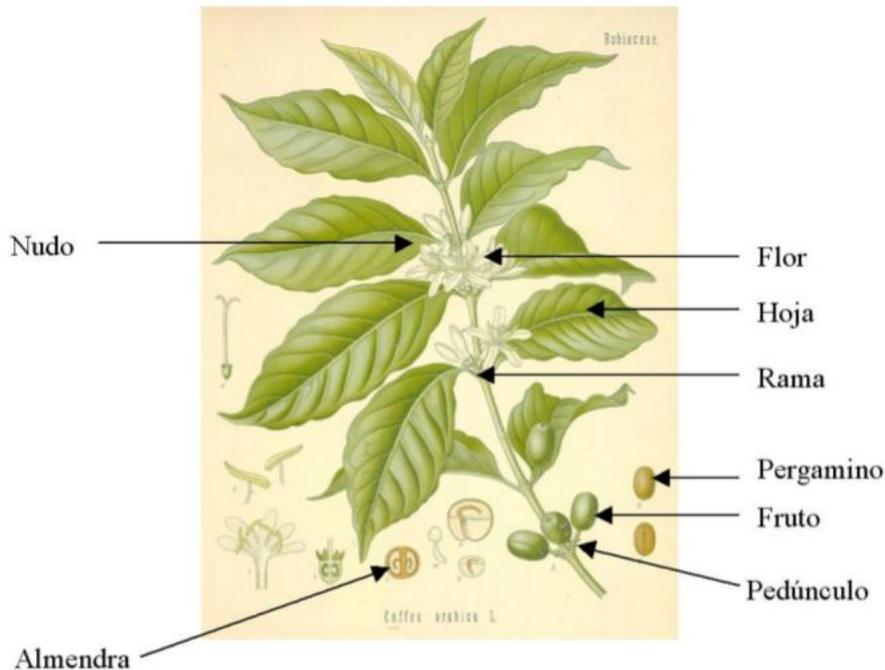


Figura 2-2: Rama de *Coffea arabica*. Imagen tomada de la base de dibujos de Cenicafé.

Existen varios tipos de estructuras vegetativas presentes en las ramas de café. Sin embargo, la presencia de algunas de ellas, como por ejemplo las flores, aparecen en cierta época del cultivo. El proceso de formación de las flores en los cafetos puede durar de 4 a 5 meses. Luego de la etapa de floración, vienen la etapa de la formación y crecimiento del fruto de café, donde las flores empiezan a desaparecer para dar paso a los frutos. Desde la antesis, que es la fase en la cual las yemas de las flores se abren, transcurren de 32 a 36 semanas para la maduración de los frutos. El tiempo de maduración depende de las condiciones ambientales en las que se encuentre el cultivo. En [10], se presentan las cuatro etapas del desarrollo de los frutos de café. La Figura 2-3 relaciona cada etapa con el peso en gramos y el tiempo en semanas. Según [10], durante el proceso de maduración los frutos de café experimentan un cambio de color, donde el espectro visible varía de 551.5nm, cuando el fruto es inmaduro, a 616,18nm para frutos maduros. Para los otros estados de maduración las longitudes de onda son cercanas a las mencionadas anteriormente.

Este trabajo identifica 6 estructuras vegetativas de la rama: tallo, hoja, flor, y frutos en

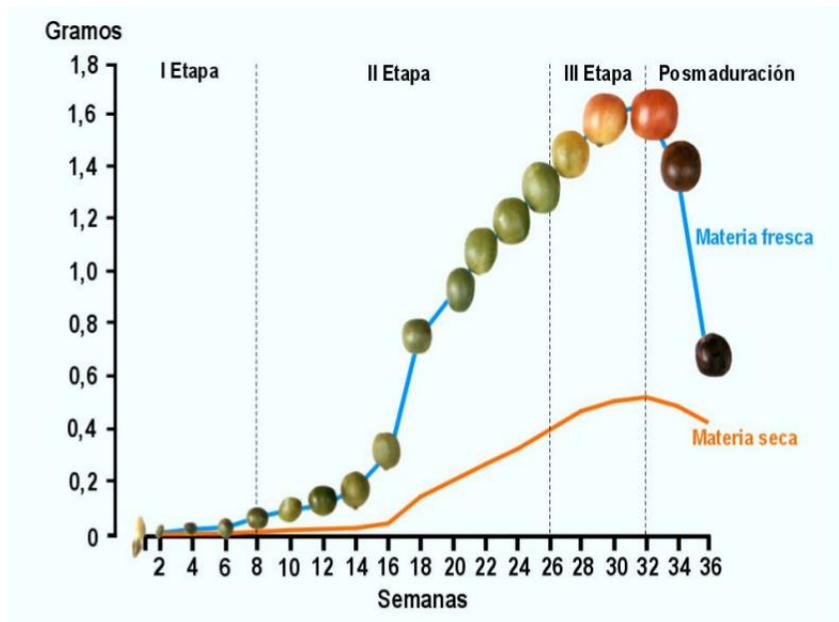


Figura 2-3: Etapas de desarrollo de los frutos de café. [1]

tres diferentes estados de maduración: inmaduro, semi-maduro y maduro. A continuación se describe brevemente cada una de las estructuras utilizadas en este trabajo, con base en la información proporcionada por café de Colombia.

Tallo:

El tallo o tronco y las ramas primarias forman el esqueleto del cafeto. La Figura 2-4 muestra el tronco y las ramas principales de un cafeto. El número de ramas nuevas es proporcional a la cosecha del cafeto, para arbustos de tres años se pueden encontrar 40 ramas en el tercio productivo.

Hoja:

Los procesos de fotosíntesis, transpiración y respiración se realizan a través de las hojas. El tiempo que duran las hojas en un cafetal es aproximadamente un año. Identificar la disminución de las hojas en las plantas ayuda al caficultor a entender que la planta está pasando por problemas nutricionales, de plagas o efectos adversos del clima.

Flor:

Las flores del cafeto aparecen en los nudos de las ramas. El proceso de formación de las flores se divide en 4 etapas: iniciación floral y diferenciación, periodo de latencia, renovación rápida del crecimiento del botón floral y apertura de las yemas. Conocer el proceso de floración



Figura 2-4: Tallo y ramas principales de un cafeto. Imagen tomada de la base de imágenes de Café de Colombia.



Figura 2-5: Hojas de un cafeto. Imagen tomada de la base de imágenes de Café de Colombia.

permite a los caficultores saber la distribución de las cosecha, estimar las necesidades de mano de obra para la futura recolección y manejar las plagas y enfermedades. La Figura 2-6 ilustra la flor del cafeto en su etapa de apertura de la yema.



Figura 2-6: Flor de cafeto. Imagen tomada de la base de imágenes de Café de Colombia.

Fruto:

De la unión del grano de polen con el óvulo se forman los frutos y las semillas del café. Como se mencionó anteriormente existen cuatro periodos o etapas de desarrollo en los frutos de café. La primera etapa va desde la fecundación hasta la 6 semana, allí el crecimiento en tamaño y peso de los frutos es lento. Luego viene la segunda etapa donde los frutos empiezan a crecer, en esta etapa se necesita de agua, de lo contrario el fruto se queda pequeño. En la tercera etapa se da una gran demanda de nutrientes, el fruto empieza a cambiar de color de verde a amarillo, en esta etapa se endurece la almendra. La falta de agua en la cuarta etapa hace que el fruto no se forme bien y se produzca el grano averanado. Finalmente, en el cuarto periodo, el cual se denomina época de maduración el fruto cambia al color rojo. Luego, si no se cosecha el fruto, se produce un fruto sobre-maduro o seco.

En este trabajo se clasificaron 3 diferentes estados de maduración, ubicados en las etapas 2, 3 y 4. El primer estado de maduración corresponde a los frutos inmaduros, el segundo a los frutos semi-maduros o pintones y el tercer estado a los frutos maduros. La Figura 2-7 ilustra los tres diferentes estados de maduración utilizados en este trabajo.

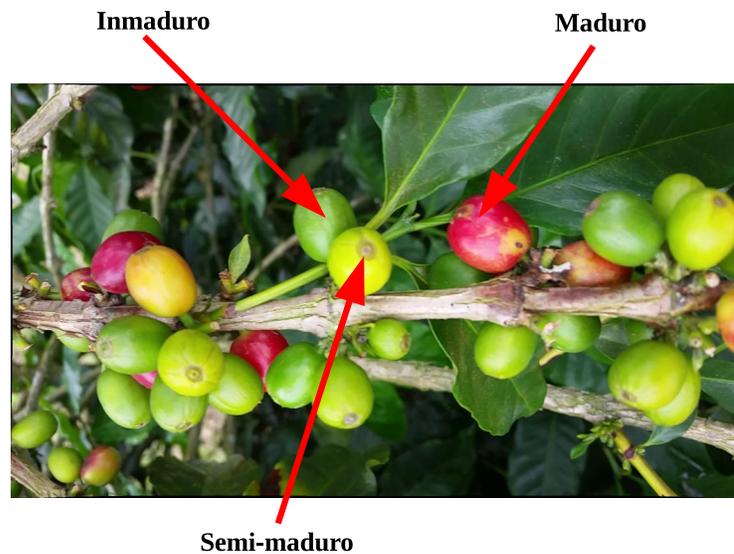


Figura 2-7: Estado de desarrollos: inmaduro, semi-maduro y maduro.

2.2. Modelo de Cámara

Los modelos de cámara son matrices con ciertas propiedades, que representan un mapeo del mundo 3D (espacio del objeto), a una imagen 2D. Modelos de cámara con centro finito e infinito son los dos grandes tipos de modelos especializados de cámara. Uno de los modelos de cámaras más simple es conocido como modelo *pinhole*. A continuación se desarrolla el modelo de cámara para sensores CCD (*Charge-Coupled Device*). Sin embargo, este modelo también se puede aplicar a otros tipos de cámaras como: imágenes de rayos X, fotografías escaneadas de negativos, etc [2].

2.2.1. Geometría proyectiva

La Figura 2-8 representa la geometría de una cámara *pinhole*. Donde, \mathbf{C} representa el centro de la cámara y \mathbf{p} el punto principal. El centro de la cámara, también conocido como centro óptico, es colocado como coordenada de origen y, adicionalmente, el plano de la imagen es ubicado al frente del centro de cámara. El eje principal es la línea que parte del centro de la cámara y es perpendicular al plano de la imagen. La distancia f que existe entre el centro de cámara \mathbf{C} y el punto principal \mathbf{p} es conocida como distancia focal. Finalmente, \mathbf{P} es el punto donde el eje principal corta el plano de la imagen.

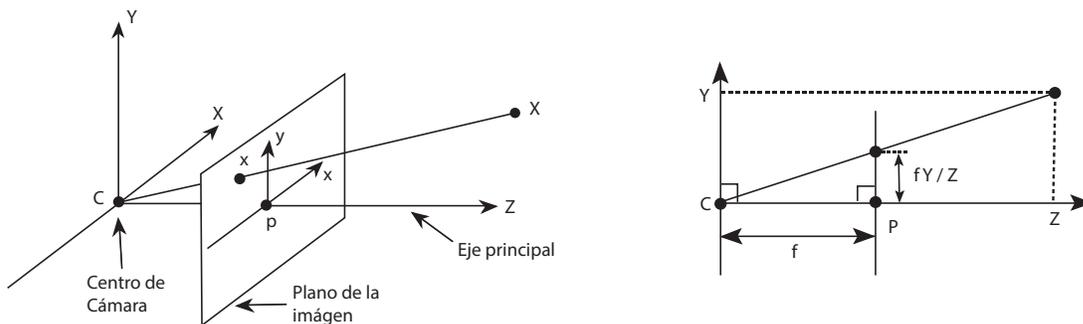


Figura 2-8: Geometría cámara *pinhole*. Imagen tomada de [2].

En el modelo de cámara *pinhole*, un punto en el espacio con coordenadas $\mathbf{X} = (X, Y, Z)^T$, es mapeado al punto en el plano de la imagen $\mathbf{x} = (x, y)^T$, donde la línea que une el punto \mathbf{X} con el centro de proyección (centro de cámara), encuentra el plano de la imagen. Por semejanza de triángulos se puede encontrar fácilmente que el punto $\mathbf{X} = (X, Y, Z)^T$, es mapeado al punto $(fX/Z, fY/Z, f)^T$, en el plano de la imagen. Sí se ignora la coordenada final se tiene que:

$$(X, Y, Z)^T \longrightarrow (fX/Z, fY/Z)^T. \quad (2-1)$$

Lo anterior describe el mapeo del espacio euclídeo \mathbb{R}^3 al espacio euclídeo \mathbb{R}^2 .

Sí el mundo real y los puntos de la imagen son representados en coordenadas homogéneas, la proyección central $(X, Y, Z)^T \rightarrow (fX/Z, fY/Z)^T$ puede ser expresada de la siguiente forma:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} fX \\ fY \\ Z \\ 1 \end{pmatrix} = \begin{bmatrix} f & & 0 \\ & f & 0 \\ & & 1 \\ & & & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \quad (2-2)$$

La Ecuación 2-2 puede ser escrita como $diag(f, f, 1) [I|0]$. Dónde $diag(f, f, 1)$ es una matriz diagonal y $[I|0]$ representa una matriz compuesta por una matriz identidad I de 3×3 más un vector columna de ceros, Ecuación 2-3.

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} fX \\ fY \\ Z \\ 1 \end{pmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \quad (2-3)$$

Con base en lo anterior, un punto en el mundo real $\mathbf{X} = (X, Y, Z, 1)^T$, puede ser representado en un punto $\mathbf{x} = (x, y, 1)^T$ de una imagen 2D de la siguiente forma:

$$\mathbf{x} = P\mathbf{X}. \quad (2-4)$$

Donde, $P = diag(f, f, 1) [I|0]$. P es conocida como la matriz de proyección de cámara.

La Ecuación 2-1, asume que el origen de las coordenadas en el plano de la imagen es el punto principal. Sin embargo, en la práctica esto no es real debido a que se presenta un *offset* entre el punto principal y el centro de la cámara, ver Figura 2-9.

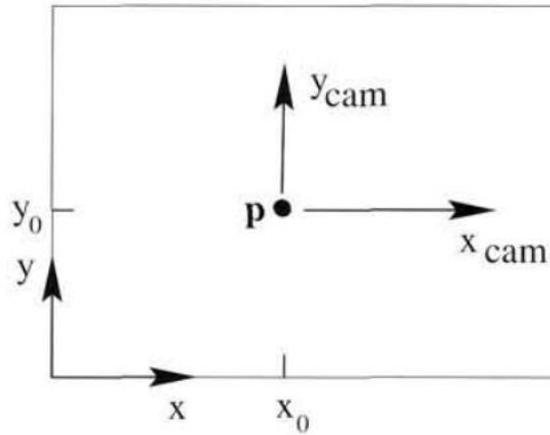


Figura 2-9: Sistema de coordenadas en la imagen (x, y) y en la cámara (x_{cam}, y_{cam}) . Imagen tomada de [2].

Debido a lo anterior la Ecuación 2-1 debe ser reescrita de la siguiente forma:

$$(X, Y, Z)^T \longrightarrow (fX/Z + p_x, fY/Z + p_y)^T \quad (2-5)$$

Donde $(p_x, p_y)^T$ representan las coordenadas del punto principal o el nuevo centro de cámara. Ahora, la Ecuación 2-2, toma la siguiente forma:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \longrightarrow \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & p_x & 0 \\ f & p_y & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \quad (2-6)$$

Ahora, escribiendo:

$$K = \begin{bmatrix} f & p_x & 0 \\ f & p_y & 0 \\ 1 & 0 & 0 \end{bmatrix}. \quad (2-7)$$

La Ecuación 2-6 se puede reescribir como:

$$\mathbf{x} = K [I|0] \mathbf{X}. \quad (2-8)$$

Donde K representa la matriz de calibración de cámara, que más adelante en este documento se tendrá que encontrar para los procesos de reconstrucción 3D, mediante procesos de calibración de cámara.

Los puntos en el espacio pueden ser expresados en términos de coordenadas euclídeas. Las dos coordenadas, las del mundo real y las de la imagen, son relacionadas mediante una rotación y translación, ver Figura 2-10.

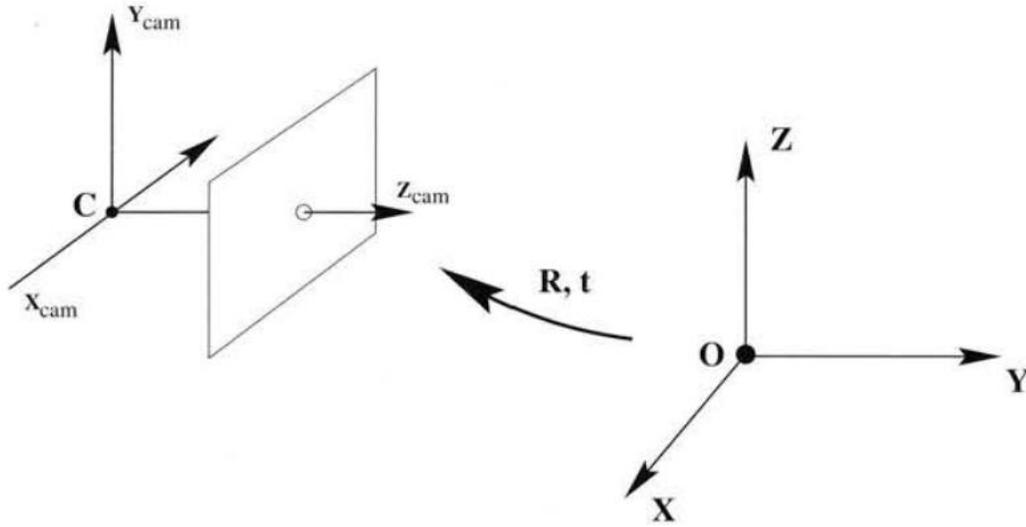


Figura 2-10: Transformación euclídea entre las coordenadas del mundo real y las coordenadas de la imagen. Imagen tomada de [2].

En la Figura 2-10, $(X_{cam}, Y_{cam}, Z_{cam})$ representan las coordenadas de un punto en el espacio proyectado en la imagen y (X, Y, Z) las coordenadas del punto en el mundo real. \mathbf{R} representa la matriz de rotación y \mathbf{t} el vector de translación. Teniendo en cuenta la rotación y translación y con base en la Ecuación 2-8, la proyección de un punto en el espacio dentro de una imagen 2D puede ser descrita de la siguiente forma:

$$\mathbf{x} = K [\mathbf{R}|\mathbf{t}] \mathbf{X}. \quad (2-9)$$

Ahora, la matriz de proyección P queda definida como:

$$P = K [\mathbf{R}|\mathbf{t}]. \quad (2-10)$$

La matriz P tiene 9 grados de libertad: 3 para la matriz K (f, p_x, p_y), 3 para \mathbf{R} y 3 para \mathbf{t} . Los parámetros que contienen la matriz K son llamados parámetros internos o intrínsecos de cámara, mientras que los parámetros de la matriz \mathbf{R} y el vector columna \mathbf{t} son conocidos

como parámetros externos o extrínsecos de cámara.

El modelo de cámara pinhole asume que las coordenadas son euclídeas y tienen la misma escala en ambas direcciones (x y y). Sin embargo, en el caso de las cámaras con sensor CCD, existe la posibilidad de tener píxeles no cuadrados, lo cual introduce factores de escala diferentes en cada dirección. Para el modelo de cámara CCD la matriz K descrita en la Ecuación 2-7 se convierte en:

$$K = \begin{bmatrix} \alpha_x & x_0 & 0 \\ & \alpha_y & y_0 & 0 \\ & & 1 & 0 \end{bmatrix}, \quad (2-11)$$

donde $\alpha_x = fm_x$ y $\alpha_y = fm_y$. Por otro lado, el centro de cámara queda determinado por $x_0 = m_x p_x$ y $y_0 = m_y p_y$. En este caso m_x y m_y son el factor de escala en cada dirección.

2.2.2. Distorsión de lentes

El modelo de cámara descrito anteriormente asume que los lentes de las cámaras son ideales y no poseen distorsión. Sin embargo, en la práctica los lentes no son perfectos debido a la manufactura de los mismos. Es más fácil construir lentes esféricos que lentes parabólicos, los cuales son matemáticamente ideales.

La distorsión radial se presenta en los píxeles cercanos a los bordes de las imágenes. Esta distorsión se produce debido a un efecto Escala: conocido como ojo de pez presente en los lentes reales. La Figura 2-11 ilustra la distorsión radial.

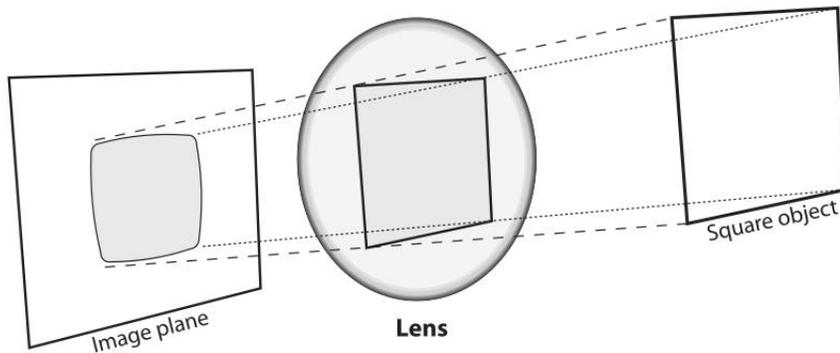


Figura 2-11: Distorsión radial. Imagen tomada de [3].

La corrección para la distorsión radial puede ser descrita de la siguiente forma:

$$x_c = x(1 + k_1r^2 + k_2r^4 + k_3r^6),$$

$$y_c = x(1 + k_1r^2 + k_2r^4 + k_3r^6), \quad (2-12)$$

dónde, (x, y) son las coordenadas originales sobre la imagen del punto distorsionado y (x_c, y_c) son la locación como resultado de la corrección.

Por su parte, otra distorsión conocida como tangencial, se presenta debido a defectos de fabricar al elaborar los lentes. Estos defectos de fabricación que hacen que el lente no este totalmente paralelo al plano de la imagen, conllevan a no tener una imagen plana, ver Figura 2-12.

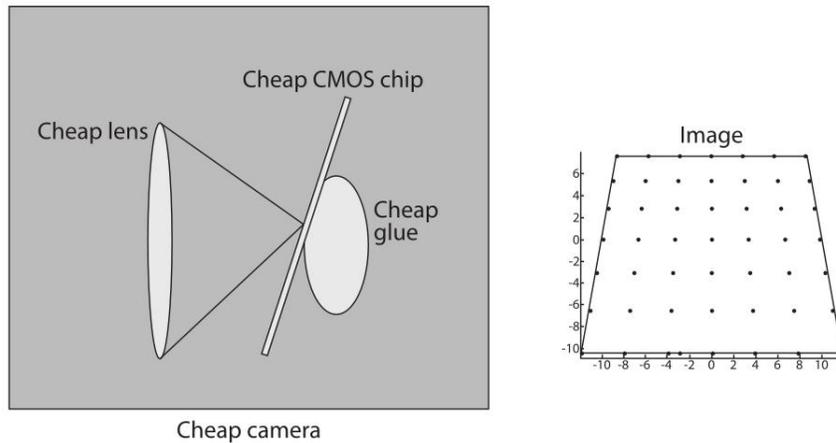


Figura 2-12: Distorsión tangencial. Imagen tomada de [3].

La distorsión tangencial está descrita en la Ecuación 2-13 [48].

$$x_c = x + [2p_1y + p_2(r^2 + 2x^2)],$$

$$y_c = y + [p_1(r^2 + 2y^2) + 2p_2x]. \quad (2-13)$$

En total existen cinco parámetros de distorsión los cuales son requeridos: k_1 , k_2 , p_1 , p_2 y k_3 .

2.3. Calibración de cámara

El proceso de calibración de cámara permite estimar los parámetros intrínsecos y extrínsecos de una cámara, los cuales son necesarios para el proceso de reconstrucción 3D. Como se menciono anteriormente, los parámetros intrínsecos son aquellos que describen el funcionamiento de la cámara: distancia focal y el centro óptico. Por su parte, los parámetros extrínsecos de la cámara son aquellos que definen la posición y la orientación del sistema de referencia de la cámara, con respecto al mundo real, es decir dan la orientación externa de la cámara.

Existen varios métodos de calibración de cámara [49]. De acuerdo al método de calibración escogido, se debe seleccionar el patrón de calibración adecuado. El método descrito por [50], estima la matriz P descrita en la Ecuación 2-10 a partir de 2 vistas diferentes de un patrón plano. Con cerca de 20 vistas se obtiene una mejor calibración. Para este método no es necesario conocer los desplazamientos del patrón entre las vistas. En este caso se utiliza un cuadro de ajedrez como patrón planar y se toman varias imágenes de este patrón desde diferentes orientaciones, ver Figura 2-13.

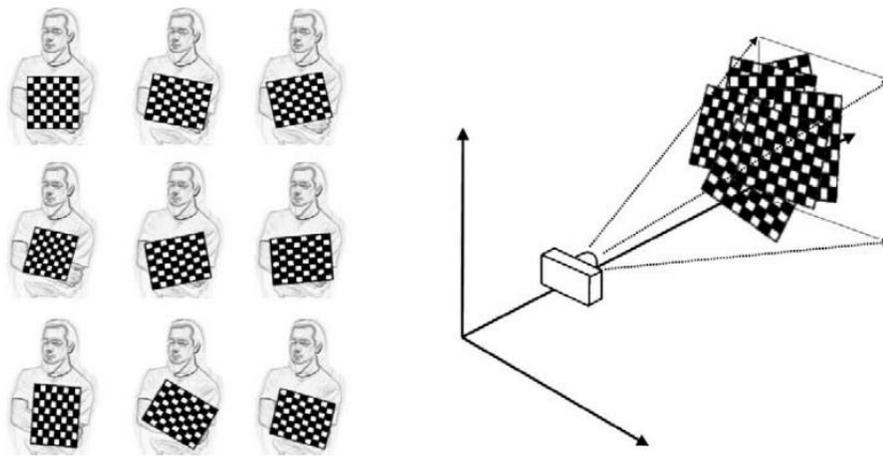


Figura 2-13: Proceso de adquisición imágenes desde diferentes orientaciones del patrón planar para el proceso de calibración de cámara. Imagen tomada de [3].

El algoritmo de calibración de cámara encuentra, en primera instancia, las esquinas de los cuadros del tablero de ajedrez, los cuales son puntos conocidos del mundo real. La transformada entre un punto en el espacio 3D (\mathbf{X}) y un punto en una imagen 2D (\mathbf{x}), está dada por la matriz de proyección P , Ecuación 2-10, cuyo tamaño es 3 x 4, tal que:

$$\mathbf{x} = P\mathbf{X}.$$

$$\begin{pmatrix} x' \\ y' \\ \omega \end{pmatrix} = \begin{pmatrix} p_{00} & p_{01} & p_{02} & p_{03} \\ p_{10} & p_{11} & p_{12} & p_{13} \\ p_{20} & p_{21} & p_{22} & p_{23} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \quad (2-14)$$

De forma general, para cualquier correspondencia $\mathbf{x}_i \longleftrightarrow \mathbf{X}_i$, se tiene $\mathbf{x}_i = P\mathbf{X}_i$, lo cual se puede expresar como:

$$\begin{pmatrix} x'_i \\ y'_i \\ \omega_i \end{pmatrix} = \begin{pmatrix} p_{00} & p_{01} & p_{02} & p_{03} \\ p_{10} & p_{11} & p_{12} & p_{13} \\ p_{20} & p_{21} & p_{22} & p_{23} \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix},$$

$$\begin{pmatrix} x'_i \\ y'_i \\ \omega_i \end{pmatrix} = (p_0 \ p_1 \ p_2 \ p_3) \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix},$$

$$\begin{pmatrix} x'_i \\ y'_i \\ \omega_i \end{pmatrix} = \begin{pmatrix} p^{0T} \\ p^{1T} \\ p^{2T} \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} = \begin{pmatrix} p^{0T} \\ p^{1T} \\ p^{2T} \end{pmatrix} \mathbf{X}_i. \quad (2-15)$$

Partiendo de que $\mathbf{x}_i = P\mathbf{X}_i$:

$$\begin{pmatrix} x_i \\ \omega_i \\ y_i \\ \omega_i \end{pmatrix} = \begin{pmatrix} p^{0T} \mathbf{X}_i \\ p^{2T} \mathbf{X}_i \\ p^{1T} \mathbf{X}_i \\ p^{2T} \mathbf{X}_i \end{pmatrix}, \quad (2-16)$$

Sí $\omega_i = 1$, de la Ecuación 2-16 se puede obtener:

$$x_i(p_{20}\mathbf{X}_i + p_{21}\mathbf{Y}_i + p_{22}\mathbf{Z}_i + p_{23}) = p_{00}\mathbf{X}_i + p_{01}\mathbf{Y}_i + p_{02}\mathbf{Z}_i + p_{03}.$$

$$y_i(p_{20}\mathbf{X}_i + p_{21}\mathbf{Y}_i + p_{22}\mathbf{Z}_i + p_{23}) = p_{10}\mathbf{X}_i + p_{11}\mathbf{Y}_i + p_{12}\mathbf{Z}_i + p_{13}. \quad (2-17)$$

De manera general la Ecuación 2-17 puede expresarse, para n puntos, de la siguiente forma:

$$A\mathbf{p} = 0, \quad (2-18)$$

donde:

$$A = \begin{pmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1 X_1 & -x_1 Y_1 & -x_1 Z_1 & -x_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1 X_1 & -y_1 Y_1 & -y_1 Z_1 & -y_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -x_n X_n & -x_n Y_n & -x_n Z_n & -x_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -y_n X_1 & -y_n Y_1 & -y_n Z_1 & -y_n \end{pmatrix},$$

$$\mathbf{p} = \begin{pmatrix} p_{00} \\ p_{01} \\ p_{02} \\ p_{03} \\ p_{10} \\ \cdot \\ \cdot \\ \cdot \\ p_{22} \\ p_{23} \end{pmatrix}.$$

A partir del cálculo de la matriz de proyección P es posible encontrar la matriz de cámara K . Sin embargo, en una proyección $\mathbf{x}_i \leftrightarrow \mathbf{X}_i$, se presenta un error en la re-proyección, conocido como error geométrico, el cual se presenta por tres fuentes: ruido en los puntos de la imagen, error en los puntos del mundo real o ruido en los dos espacios (2D y 3D). La Figura 2-14 ilustra un ejemplo del error geométrico los puntos rojos ilustran la verdadera ubicación de los puntos.

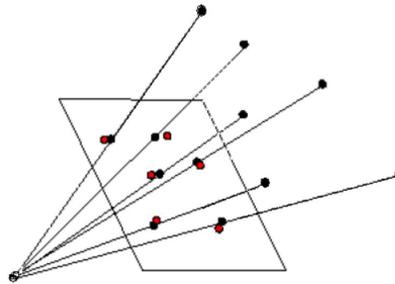


Figura 2-14: Error geométrico presente en el proceso de calibración de cámara.

El proceso de calibración de cámara debe ahora calcular P , basado en los puntos medidos en la imagen \mathbf{x}_i y en las proyecciones $\hat{\mathbf{x}}_i$ de los puntos \mathbf{X}_i . El error geométrico local, es la distancia euclídea entre los puntos \mathbf{x}_i y $\hat{\mathbf{x}}_i$. Para todo el conjunto de puntos, el error geométrico

será la suma de las distancias euclídeas de todos los puntos (n):

$$\sum_{i=1}^n d_{eucli}(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 = \sum_{i=1}^n d_{eucli}(\mathbf{x}_i, P\mathbf{X}_i)^2. \quad (2-19)$$

Sí las medidas del error geométrico tienen una distribución gaussiana, la solución de minimizar el error de la siguiente forma:

$$\min_P = \sum_{i=1}^n d_{eucli}(\mathbf{x}_i, P\mathbf{X}_i)^2. \quad (2-20)$$

Lo anterior representa la estimación de máxima probabilidad de P . La minimización del error geométrico requiere el uso de técnicas de optimización iterativas, tales como Levenberg-Marquardt [51].

2.4. Estructura a partir de movimiento

La estructura a partir de movimiento o (SfM) por sus siglas en inglés *Structure from Motion*, es un tópico importante en el area de la visión por computador. Por medio de esta técnica es posible realizar una reconstrucción 3D de un objeto en el espacio, a partir de múltiples vistas tomadas alrededor del objeto, ver Figura 2-15.

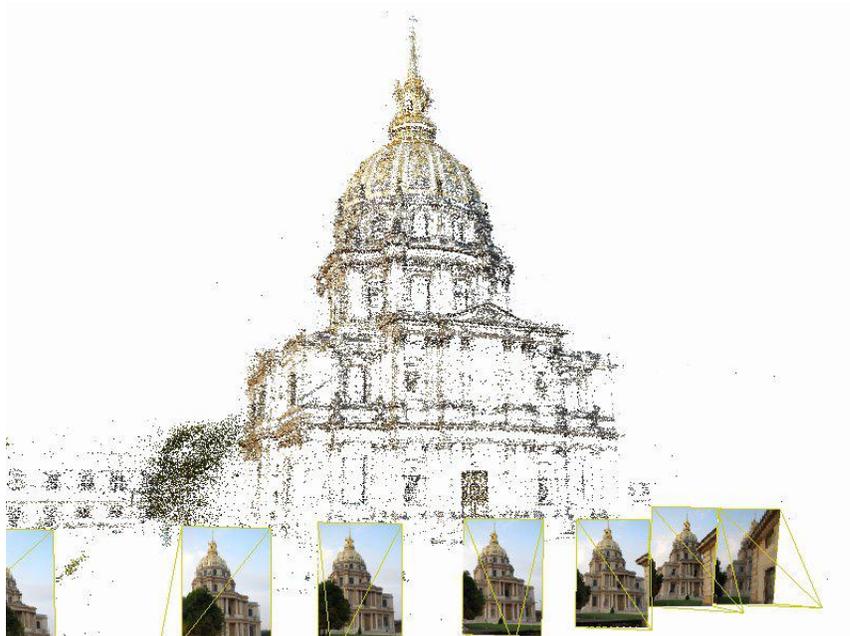


Figura 2-15: Estructura a partir de movimiento. Tomado de [4]

El proceso de reconstrucción 3D utilizando el método de estructura a partir de movimiento comienza extrayendo características en todas las imágenes capturadas del objeto. Luego se encuentran las correspondencias entre los puntos característicos vista a vista, utilizando técnicas como flujo óptico o aproximación del vecino más cercano. Después de encontrar la correspondencia entre las características se procede a determinar el movimiento de la cámara entre cada vista, matrices de rotación \mathbf{R} y translación \mathbf{t} , a través del cálculo de las matrices Fundamental \mathbf{F} y Esencial \mathbf{E} . Una vez encontrado el movimiento de cada vista se proyectan los puntos característicos encontrados mediante un proceso de triangulación, formando así una estructura 3D del objeto.

2.4.1. Extracción de puntos característicos utilizando SIFT

Los puntos característicos en la vistas son encontrados utilizando el algoritmo conocido como (SIFT) por sus siglas en inglés *Scale-Invariant Feature Transform* [5]. El algoritmo de SIFT encuentra características invariantes a la escala y rotación de la imagen. Adicionalmente, el método de SIFT es robusto a las distorsiones afines, cambios en puntos de vistas 3D, adición de ruido y cambios en la iluminación. Los anteriores son problemas presentes al trabajar con imágenes de campo como las que se utilizaron en este trabajo.

Los principales pasos para la generación de puntos característicos (*keypoints*) utilizados por SIFT son:

1. El primer paso consiste en identificar ubicaciones y escalas que pueden ser asignadas repetidamente bajo diferentes vistas del mismo objeto. Para esto se crean varias escalas de la imagen y luego se calcula la diferencia entre las mismas, ver Figura 2-16.

La escala de espacio de una imagen esta definida como una función, $L(x, y, \sigma)$, que es producida de la convolución de una función gaussiana variable a la escala, $G(x, y, \sigma)$, con una imagen, $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y),$$

dónde $*$ es la convolución entre $I(x, y)$ y

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \quad (2-21)$$

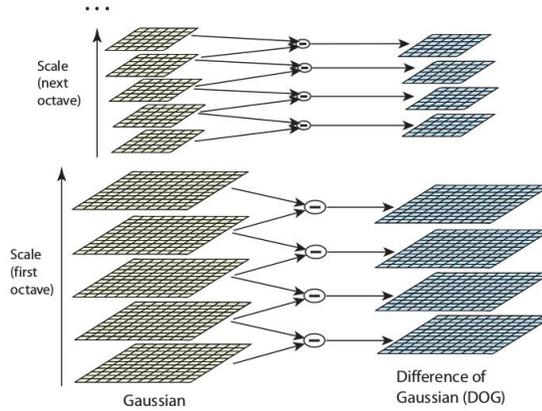


Figura 2-16: Diferencia de Gaussiano en diferentes escalas de la imagen. Tomado de [5]

La detección de la ubicación de un punto característico en diferentes escalas se realiza mediante el cálculo de la diferencia gaussiana o (DoG), $D(x, y, \sigma)$, la cual se obtiene de la diferencia entre dos escalas adyacentes.

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned} \quad (2-22)$$

En la Figura 2-16 por cada octava de espacio de escala, la imagen inicial es repetidamente convolucionada con la función gaussiana de la Ecuación 2-21, para producir el conjunto de imágenes en diferente espacio de escala. Luego, las imágenes son restadas para producir las imágenes de diferencia gaussiana. Después, en cada octava, la imagen gaussiana es sub-muestreada por un factor de 2 y el proceso se repite.

2. El siguiente paso consiste en asignar la orientación de los puntos clave. Para asignar una orientación consistente para cada punto clave el algoritmo de SIFT se basa en las propiedades locales de la imagen. Por cada imagen muestreada $L(x, y)$, la magnitud del gradiente $m(x, y)$, y orientación $\theta(x, y)$, es precalculado usando diferencia de píxeles de la siguiente forma:

$$\begin{aligned} m(x, y) &= \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}. \\ \theta(x, y) &= \tan^{-1}(L(x+1, y) - L(x-1, y)) / (L(x, y+1) - L(x, y-1)). \end{aligned} \quad (2-23)$$

Un histograma de orientación con 36 bins es formado desde las orientaciones de los gradientes de puntos de muestra dentro de una región alrededor de un punto clave. Los picos de los histogramas son detectados y corresponden a las direcciones de los gradientes locales. Los picos altos son utilizados para identificar un punto clave dentro de la imagen.

Al finalizar este proceso se tendrán, por cada imagen utilizada para la reconstrucción, varios puntos característicos.

2.4.2. Emparejamiento de puntos característicos utilizando Flujo Óptico

Como se menciona anteriormente, existen diferentes técnicas para encontrar la correspondencia entre puntos característicos entre frame y frame, por ejemplo: emparejamiento por fuerza bruta, aproximación del vecino más cercano y flujo óptico. Si se desea aplicar flujo óptico para el proceso de emparejamiento de puntos clave desde una imagen a otra, se debe asumir que ambas imágenes son parte de una secuencia y una relativamente cercana a la otra, como es el caso de los frames de un vídeo. Muchos métodos que calculan el flujo óptico comparan una pequeña región, conocida como ventana de búsqueda, alrededor de cada punto desde una imagen A a la misma área en una imagen B. Siguiendo una regla común en la visión por computador, llamada restricción constante del brillo, la ventana de búsqueda en una imagen no debe cambiar drásticamente en la otra, por lo tanto el valor absoluto de la resta de ambas debe ser cercano a cero.

En caso de que la intensidad en un píxel, perteneciente a un objeto, no cambie entre dos frames consecutivos, y que los píxeles vecinos tengan un movimiento similar, si un píxel $I(x, y, t)$, en el primer frame se mueve una distancia (dx, dy) en el siguiente frame, después de un (dt) , la intensidad de estos píxeles es la misma y se puede decir que:

$$I(x, y, t) = I(x + dx, y + dy, t + dt). \quad (2-24)$$

Luego de una aproximación por series de Taylor del lado derecho, removiendo términos iguales y dividiendo en dt se obtiene la siguiente ecuación:

$$f_x u + f_y v + f_t = 0 \quad (2-25)$$

dónde,

$$f_x = \frac{\partial f}{\partial x}; f_y = \frac{\partial f}{\partial y}; u = \frac{dx}{dt}; v = \frac{dy}{dt} \quad (2-26)$$

La Ecuación 2-25 es conocida como la Ecuación de Flujo Óptico. f_x y f_y son los gradientes de la imagen, mientras que f_t es el gradiente a lo largo del tiempo. Las incógnitas entonces son

(u, v) . Esta única Ecuación sólo se puede resolver con dos incógnitas. Existen varios métodos para resolver este problema como el método de Lukas-Kanade [52].

El método de Lucas-Kanade toma una ventana de 3×3 alrededor del punto, se asume que todos los nueve puntos tienen el mismo movimiento. Ahora se tiene que encontrar (f_x, f_y, f_t) para estos nueve puntos. El problema se convierte entonces en resolver 9 ecuaciones con dos variables desconocidas. Una solución se obtiene al aplicar un ajuste de mínimos cuadrados, Ecuación 2-27.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i f_{x_i}^2 & \sum_i f_{x_i} f_{y_i} \\ \sum_i f_{x_i} f_{y_i} & \sum_i f_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i f_{x_i} f_{t_i} \\ -\sum_i f_{y_i} f_{t_i} \end{bmatrix} \quad (2-27)$$

Lo anterior funciona para movimientos pequeños presentes en las secuencias de imágenes. Sin embargo, se pueden presentar movimientos más grandes, entonces utilizar lo descrito anteriormente generaría errores encontrando el movimiento de los puntos clave. Para solucionar esto se trabaja con pirámides donde los movimientos pequeños son eliminados y los movimientos grandes se convierten en pequeños, permitiendo así aplicar el método de Lucas-Kanade a escala.

La Figura 2-17 ilustra el resultado de encontrar puntos característicos en dos imágenes consecutivas (A y B) y luego emparejarlos utilizando el algoritmo de flujo óptico.

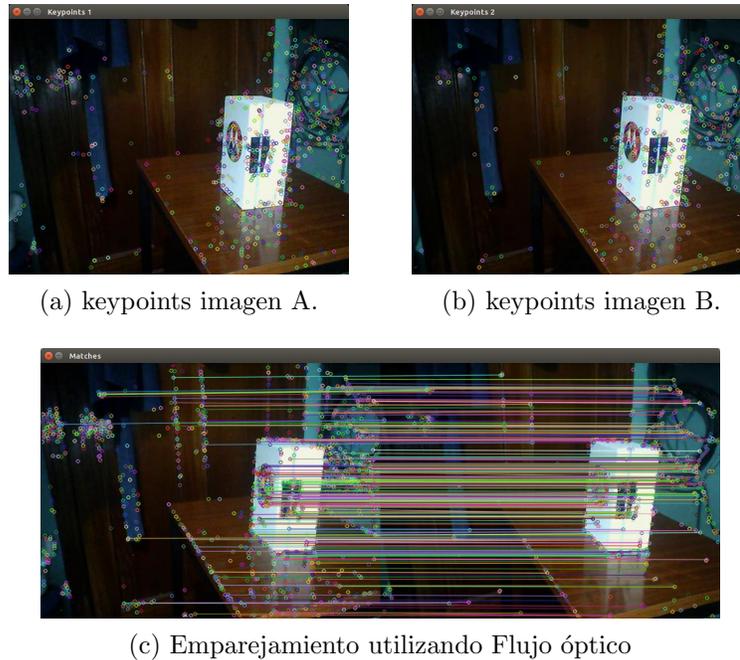


Figura 2-17: Extracción de *keypoints* utilizando SIFT y emparejamiento por medio de flujo óptico

2.4.3. Cálculo de la matriz fundamental

Después de obtener la correspondencia de los *keypoints* de todas las vistas, se procede a calcular la matriz fundamental (\mathbf{F}) por cada movimiento. La matriz fundamental (\mathbf{F}) es una representación algebraica de la geometría epipolar. Para encontrar la matriz fundamental en cada movimiento es necesario tener en cuenta lo siguiente:

- Dado un par de imágenes, para cada punto x en una imagen, existe una línea epipolar correspondiente l , en la otra imagen.
- Dado un conjunto de correspondencias $x_i \leftrightarrow x'_i$ en dos imágenes, la matriz fundamental F satisface la condición $x'_i F x_i = 0$, para todo i .
- Con x_i y x'_i conocidos, la anterior Ecuación es lineal con respecto a los parámetros de F .

El número de puntos característicos que se usen para la estimación de la matriz fundamental determina el método con el cual se debe calcular. Si por ejemplo se tienen 8 correspondencias se puede obtener la matriz (\mathbf{F}) con una solución lineal, si se tienen sólo 7 correspondencias se tiene que utilizar una solución no lineal y, si por lo contrario, se tienen más de 8 correspondencias, es posible obtener una solución usando mínimos cuadrados o RANSAC.

2.4.4. Cálculo del movimiento de la cámara

Luego de calcular las matrices fundamentales \mathbf{F}_j en cada movimiento se procede a encontrar la rotación y translación de la cámara entre cada vista. Para lo anterior es necesario calcular la matriz esencial \mathbf{E} , la cual está descrita por la Ecuación 2-28.

$$\mathbf{E} = \mathbf{K}^T \mathbf{F} \mathbf{K}, \quad (2-28)$$

dónde \mathbf{K} es la matriz de parámetros internos de la cámara encontrada luego de realizar el proceso de calibración de cámara. \mathbf{F} es la matriz fundamental encontrada anteriormente. La matriz esencial \mathbf{E} puede ser definida como $\mathbf{E} = [\mathbf{t}]_x \mathbf{R}$. La rotación y translación puede ser calculada a través de la descomposición por valores singulares de la matriz esencial, tal como se muestra a continuación:

$$\text{SVD}(\mathbf{E}) = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T, \quad (2-29)$$

dónde \mathbf{U} y \mathbf{V} son matrices ortogonales de 3×3 y $\mathbf{\Sigma}$ es una matriz diagonal de 3×3 que contiene en su diagonal dos números idénticos y un valor de cero. Para calcular la rotación y translación es necesario definir la siguiente matriz anti-simétrica:

$$\mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{W}^{-1} = \mathbf{W}^T = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-30)$$

La translación está dada por:

$$[\mathbf{t}]_x = \mathbf{U}\mathbf{W}\mathbf{\Sigma}\mathbf{U}^T. \quad (2-31)$$

Y la rotación:

$$\mathbf{R} = \mathbf{U}\mathbf{W}^T\mathbf{V}^T. \quad (2-32)$$

2.4.5. Triangulación

Una vez se han encontrado las matrices de rotación y la translación entre cada vista, se procede a proyectar los puntos en el espacio. La Figura 2-18 ilustra el proceso de triangulación de dos puntos.

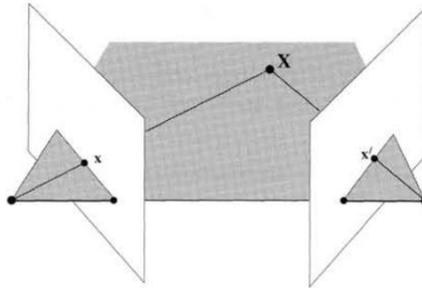


Figura 2-18: Triangulación de puntos. Tomado de [2]

Dados dos puntos correspondientes (x y x') en dos imágenes tomadas con las misma cámara, es posible calcular la ubicación (X , Y y Z) en el espacio. Al realizar las siguientes operaciones:

En cada imagen se conoce que:

$$x = KPX \quad x' = KP'X, \quad (2-33)$$

Si se multiplican ambas ecuaciones por K^{-1} se obtiene:

$$K^{-1}x' = K^{-1}KP'X, \quad K^{-1}x = K^{-1}KPX \quad (2-34)$$

Como $K^{-1}K = I$, la Ecuación 2-34 se convierte en:

$$K^{-1}x' = P'X \quad K^{-1}x = PX \quad (2-35)$$

El anterior sistema de ecuaciones pueden ser combinado en la forma:

$$AX = 0. \quad (2-36)$$

Finalmente, A es:

$$A = \begin{bmatrix} xp^{3T} - p^{1T} \\ yp^{3T} - p^{2T} \\ x'p'^{3T} - p'^{1T} \\ y'p'^{3T} - p'^{2T} \end{bmatrix}. \quad (2-37)$$

Con base en lo anterior se puede resolver el problema con una solución lineal y encontrar los puntos (X , Y y Z) en el espacio.

2.5. Máquinas de vectores de soporte

Las máquinas de vectores de soporte o (SVM) por sus siglas inglés *Support Vector Machine* son una técnica de clasificación y regresión desarrollada por Cortes y Vapnik en 1995 [53]. Esta técnica de clasificación busca generar hiper-planos que permitan separar las clases lo mejor posible aún si es necesario proyectar los datos en un espacio diferente al original. Aunque existen varios hiper-planos que permiten separar las clases, algunos poseen mayor margen que otros como se ilustra en la Figura 2-19.

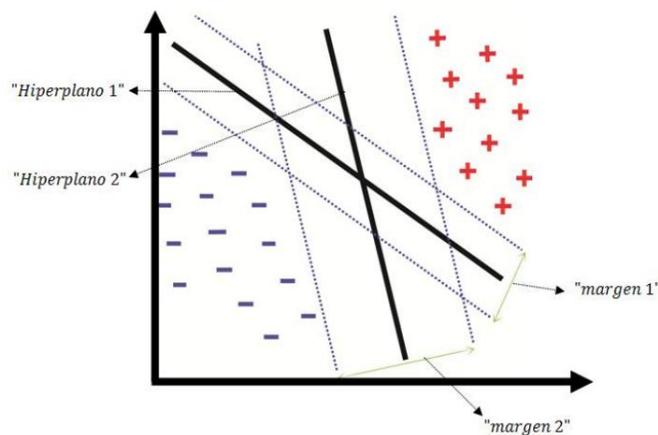


Figura 2-19: Ejemplo de varios hiper-planos de separación con sus respectivos márgenes para una SVM.

Las SVM buscan maximizar el margen de separación y minimizar el error separando las clases. En caso de que las clases no sean separables linealmente se aplica una transformación no lineal $\Phi(x)$ conocida como Kernel el cual proyecta los datos de entrada a un espacio mayor. El modelo estándar de una SVM esta definido como:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (2-38)$$

Sujeto a:

$$y_i(\langle \mathbf{W}, \phi(\mathbf{x}_i) \rangle + b) + \xi_i \geq 1, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, l$$

dónde \mathbf{x}_i son los vectores de entrenamiento, sujeto a $\mathbf{x}_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}^l$ tal que $y_i \in \{0, 1, 2, \dots, m\}$. (m) es número de clases y (l) el número de muestras de entrenamiento usadas. $\phi(\mathbf{x}_i)$ mapea \mathbf{x}_i a un espacio de dimensión mayor y $C > 0$ es el parámetro de regularización. El modelo dual de la Ecuación 2-38 esta definido:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j k(x_i, x_j) - \sum_{i=1}^l \alpha_i \quad (2-39)$$

Sujeto a:

$$\sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l$$

$K(x_i, x_j) \equiv \Phi(x_i)^T \Phi(x_j)$ es la función de kernel, algunas funciones de kernel utilizadas en las SVM son: lineal, polinomial, funciones de base radial o RBF (*Radial Basis Function*) y sigmoidea. El kernel gaussiano está definido como:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad (2-40)$$

con $\gamma > 0$.

La Figura **2-20** ilustra un ejemplo de la arquitectura de una SVM, donde la función de kernel $K(x_i, x_j)$ es escogida a priori y los parámetros de pesos λ_n y el umbral b son encontrados durante la fase de entrenamiento.

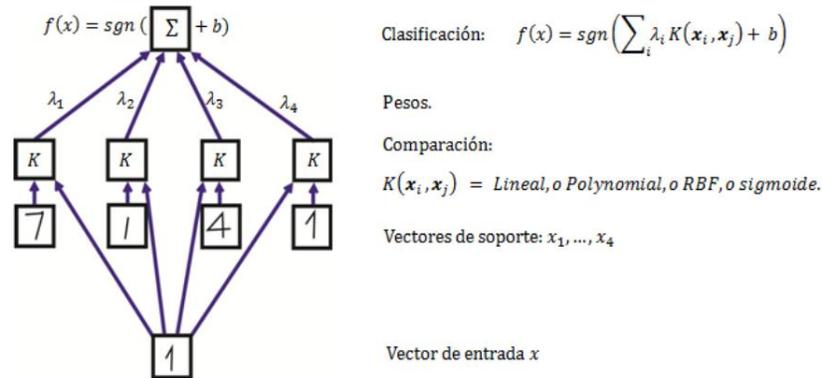


Figura 2-20: Ejemplo arquitectura de una SVM. Adaptada de [6]

La primera capa de pesos en la Figura 2-20 son un sub-conjunto de datos de entrenamiento conocidos como vectores de soporte, y la segunda capa de pesos $\lambda_i = y_i \alpha_i$ son calculados por medio de los multiplicadores de Lagrange. Finalmente la clasificación se realiza de acuerdo al signo de la función de decisión:

$$f(x) = \text{sgn}\left(\sum_i \lambda_i K(x_i, x_j) + b\right) \quad (2-41)$$

Las SVM son esencialmente clasificadores binarios. Sin embargo, pueden adoptar tareas de clasificación múltiple (multi-clase), adoptando las técnicas conocidas como: "1 vs 1z el "1 vs todos"[54].

3 Clasificación de las estructuras vegetativas en ramas de café

El sistema propuesto está compuesto por 5 fases: reconstrucción 3D, filtrado, sub-muestreo, extracción de características y clasificación, dicho sistema se ilustra en la Figura **3-1**. Los datos de entrada fueron secuencias de imágenes extraídas de vídeos tomados en campo con un dispositivo móvil, los demás procesos del sistema fueron realizados de manera *offline*. En total se adquirieron 12 vídeos de ramas de café. Se seleccionaron los *frames* que entregaron mejor información por cada rama, aproximadamente 90 por vídeo. Cada rama fue reconstruida usando las técnicas de *Structure from Motion* (SfM) y *Patch-Multi-View Stereo* (PMVS) [55, 56, 7], obteniendo una nube de puntos por cada rama.

Cada nube de puntos fue filtrada, removiendo en primera instancia *outliers* generados en la fase reconstrucción 3D. Posteriormente los puntos que se encontraban a mayor profundidad fueron considerados fondo y se eliminaron por medio de un filtro pasa-banda. Con la rama filtrada, se realizó un proceso de sub-muestreo utilizando filtrado por voxelización, donde el tamaño del vóxel fue seleccionado automáticamente dependiendo de la geometría de la rama. Para cada vóxel, se extrajo información de diferentes espacios de color (Características 2D). Adicionalmente, fueron obtenidas características 3D a partir de las normales calculadas sobre regiones de cada nube de puntos sub-muestreada. Con la nube de puntos sub-muestreada y las imágenes anotadas manualmente, se obtuvo una nube de puntos anotada (*GroundTruth*) por cada rama, estas nubes de puntos *GroundTruth* fueron utilizadas para los procesos de entrenamiento y validación del sistema. Finalmente, con la información 2D y 3D obtenida en 8 ramas se entrenó una SVM para clasificar seis estructuras vegetativas: tallo, hojas, flores, frutos inmaduros, semi-maduros y maduros. Las 4 ramas restantes se utilizaron para evaluar el desempeño del método propuesto.

Los datos de entrada al sistema propuesto provienen de los videos capturados en campo por el dispositivo móvil. Sin embargo, el procesamiento de la información se realiza de manera *offline* en una estación de trabajo como por ejemplo un computador o un servidor. El costo computacional del sistema desarrollado varía según el número de frames procesados que depende de la longitud de cada rama. El tiempo empleado por el sistema para entregar una rama clasificada fue en promedio de 40 a 60 minutos, los procesos de reconstrucción 3D y extracción de características son los que requieren de mayor tiempo para su ejecución.

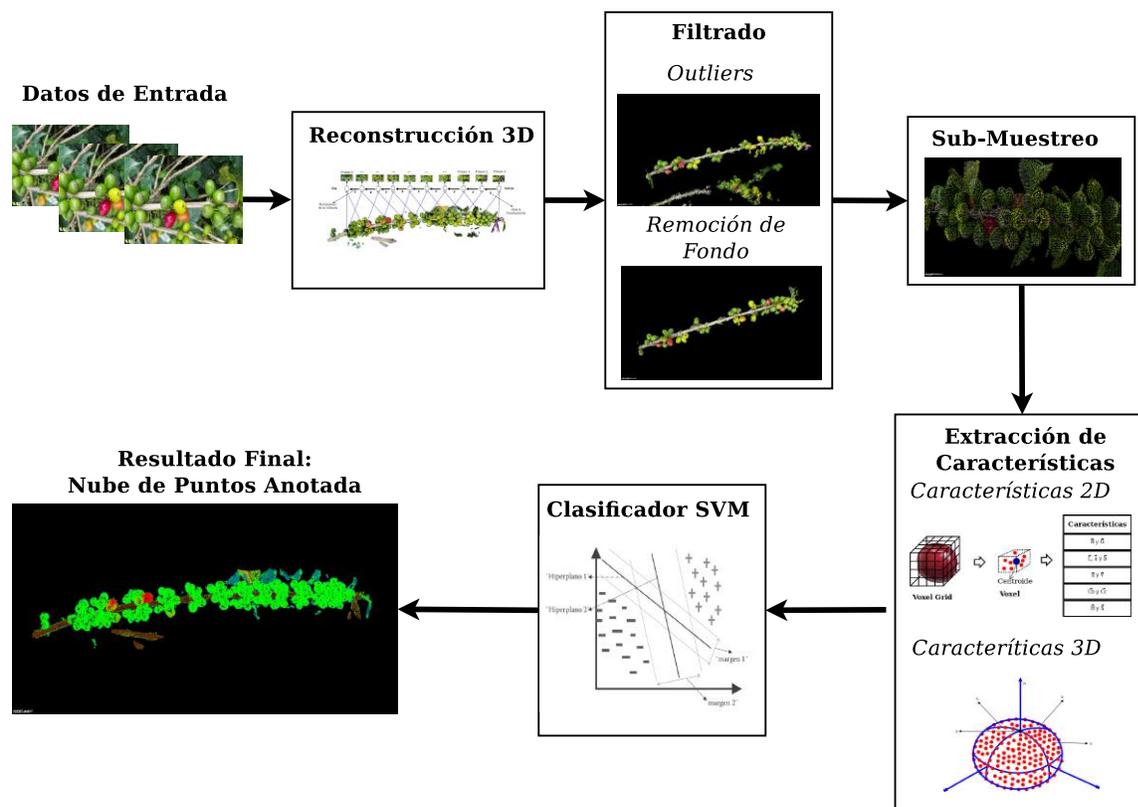


Figura 3-1: Sistema propuesto para clasificación de estructuras vegetativas presentes en ramas de café.

3.1. Datos de entrada

3.1.1. Adquisición de las imágenes

La adquisición de las imágenes se realizó sobre un lote de tercera cosecha en la Estación Experimental Naranjal, ubicada en Chinchina - Caldas por el personal de Cenicafé. La variedad de café utilizada corresponde a *Coffea arabica* variedad castillo®. Las ramas con las que se trabajó se encontraban entre el cruce 6 y 12. Las imágenes usadas fueron adquiridas con una cámara posterior integrada en un dispositivo móvil, Samsung S5 SM-G900M, ubicada a una distancia de 8 y 15 centímetros de la rama, por medio de un sujetador, como se ilustra en la Figura 3-2a. El sujetador creado contaba con un botón para el inicio y finalización de la captura de vídeo, ver Figura 3-2b. La cámara fue configurada para adquirir vídeos en Full-HD (1920 × 1080), 30fps, sin activación del flash. Se capturaron vídeos sobre el tercio productivo del arbusto a lo largo de cada rama. La información fue adquirida bajo condiciones de campo no controladas en diferentes horas del día (mañana y tarde) con luz directa y difusa con el objetivo de garantizar diferentes ambientes de iluminación. Los vídeos fueron tomados durante una temporada de cosecha. En total, se realizaron 8 pases en el segundo trimestre del año 2015. La iluminación fue la del medio ambiente presente en cada adquisi-



(a) Soporte y sujetador del dispositivo móvil para adquirir información en campo. (b) Botón de inicio/finalización de captura de vídeos.

Figura 3-2: Ejemplo de adquisición de imágenes en un cafeto.

ción de información y el fondo de las imágenes estaba conformado por suelo, malas hierbas, hojas secas, y ramas de otros árboles.

3.1.2. Anotación manual de las imágenes capturadas en campo

El proceso de anotación manual de las imágenes se realizó con la herramienta desarrollada en [57], con la que a partir de marcas sencillas se genera automáticamente una Región de Interés (ROI), la cual se etiqueta con la estructura vegetativa correspondiente. La herramienta es capaz de analizar secuencias de vídeo de manera rápida y con tiempo mínimo para el proceso de anotación. La Figura 3-3 muestra dos de las imágenes utilizadas junto con su respectiva imagen anotada. Las estructuras vegetativas presentes en la escena se etiquetan utilizando diferentes colores. Por ejemplo, los tallos son representados con un color café, las hojas con un verde azulado, las flores con un verde oscuro, los frutos inmaduros con un verde claro, los semi-maduros con un amarillo, los maduros con un rojo, y el fondo con un gris claro.

3.1.3. Generación de la Nube de puntos Ground Truth

Luego de obtener la nube de puntos sub-muestreada explicada en la Sección 3.4, se procedió a reconstruir una nube de puntos anotada manualmente (Ground Truth), donde cada punto corresponde a una de las seis clases: tallo, hoja, flor, fruto inmaduro, fruto semi-maduro y fruto maduro. Para generar la nube de puntos *grountruth*, por cada punto p_i de la nube de puntos sub-muestreada Q , se identificó las imágenes 2D de las cuales provino el punto, luego utilizando la Ecuación 3-1, se calculó la posición (u, v) del píxel p . Utilizando la ubicación

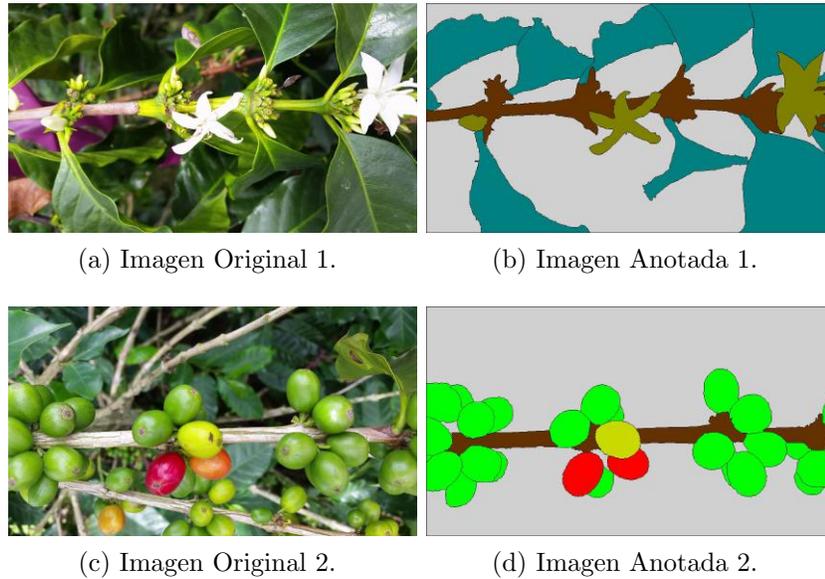


Figura 3-3: Imágenes adquiridas junto con su respectiva anotación.

del píxel p sobre las imágenes anotadas manualmente se extrae la información de color de cada imagen y se formó un arreglo llamado \mathbf{C} , donde se almacena la información de color de los píxeles presentes en las imágenes, de los cuales se proyectó el punto p_i . Finalmente, se asignó el valor de la mediana de cada espacio de color RGB del arreglo \mathbf{C} , a las componentes de color de la nube de puntos Groundtruth en el punto p_i . Se utilizó la mediana para no generar nuevos datos de color, ya que lo que se buscaba era generar una nube de puntos Groundtruth que mantuviera las etiquetas (colores), de las imágenes anotadas. La ubicación (u,v) de cada píxel se realizó utilizando la siguiente ecuación:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R|t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (3-1)$$

donde, X , Y , y Z son las coordenadas del punto p_i , K es la matriz de parámetros intrínsecos encontrada en la fase de calibración de cámara, R la matriz de rotación y t el vector de traslación. R y t son calculados en la fase de reconstrucción.

La Figura 3-4 muestra dos de las nubes de puntos Ground Truth generadas, a los frutos maduros se les asignó color rojo, al tallo color café, a las hojas color verde azulado, a las flores color verde oscuro, a los frutos inmaduros color verde claro y a los frutos semi-maduros color amarillo.

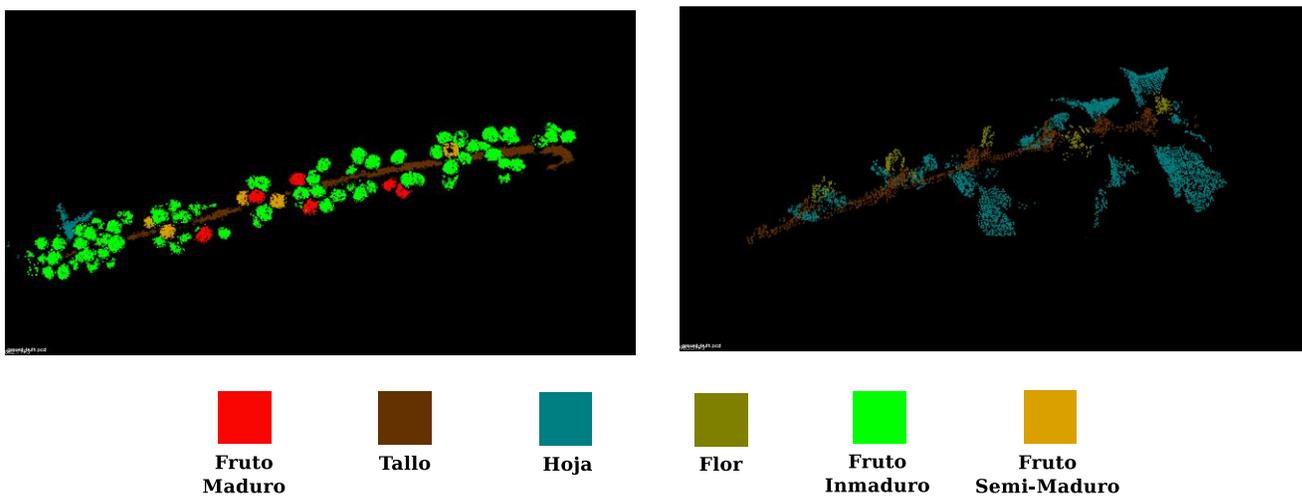


Figura 3-4: Ejemplo nubes de puntos GroundTruth generadas.

3.2. Reconstrucción 3D

El primer paso en el proceso de reconstrucción 3D consiste en encontrar los parámetros intrínsecos de la cámara utilizada, es decir la matriz K , ecuación 2-11. Para esto, se usó el algoritmo de calibración de cámara desarrollado en [50], donde se requieren pocas imágenes de un patrón planar desde diferentes orientaciones. Se utilizó un patrón planar de un cuadro de ajedrez con un ancho y alto de 1,7cm por cada cuadro, la Figura 3-5 ilustra el patrón utilizado. La Tabla 3-1 muestra los parámetros intrínsecos encontrados después de la fase de calibración.

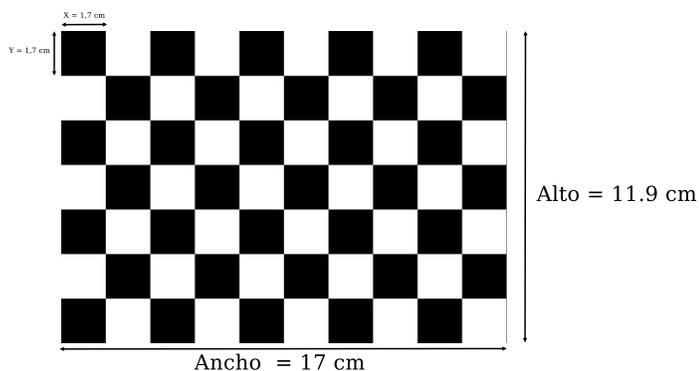
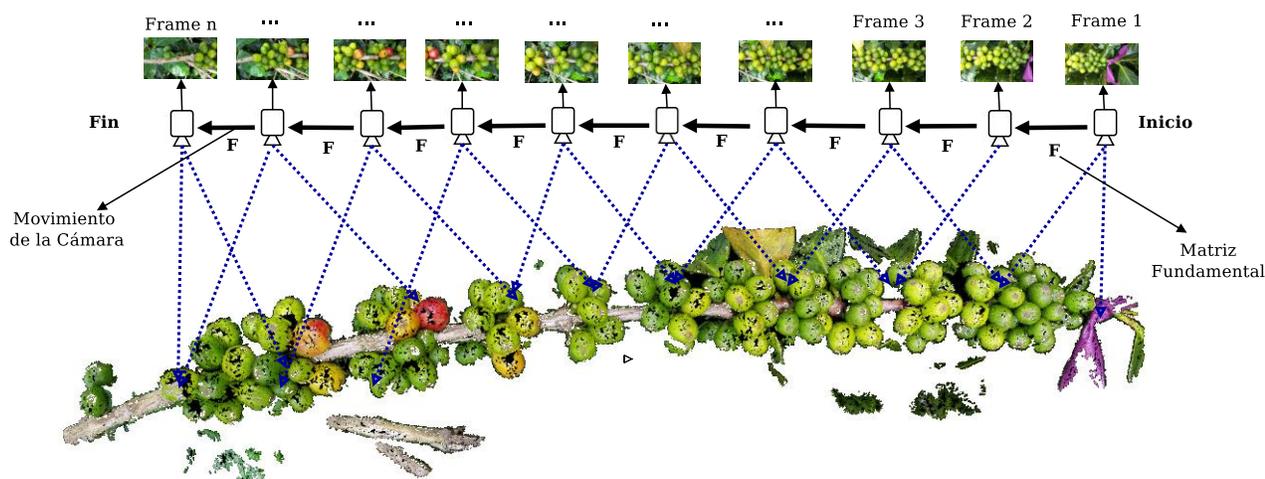


Figura 3-5: Patrón utilizado para la fase de calibración de cámara.

Tabla 3-1: Parámetros intrínsecos de la cámara.

Distancia Focal		Centro de Cámara		Coeficientes de Distorsión				
f_x	f_y	C_x	C_y	k_1	k_2	k_3	k_4	k_5
698.1	698.1	319.5	179.5	711,2	87,8	0	0	-3.58

La Figura 3-6, ilustra el proceso usado para la implementación de SfM. La cámara del dispositivo móvil captura vídeos en alta resolución a lo largo de cada rama iniciando en su extremo y recorriendo la rama hacia el tronco central, ver Figura 3-2b. La longitud por rama oscila entre los 40 y 60 centímetros y la duración promedio de los vídeos es de 15 segundos. En cada vídeo se selecciona uno de cada cinco *frames*. El número de imágenes utilizadas para la reconstrucción de cada rama es en promedio de 90.

**Figura 3-6:** Proceso usado para la implementación de SfM.

La reconstrucción 3D de cada rama se realizó implementando el enfoque de Bundler SfM utilizando las librerías desarrolladas por [55, 56, 7]. Bundler SfM es un sistema de visión de máquina que implementa SfM para la reconstrucción 3D de escenas provenientes de colecciones desordenadas de imágenes, por ejemplo fotos de Internet. A continuación se ilustran los resultados de las fases principales para obtener una estructura 3D de una escena utilizando Bundler SfM.

En primer lugar Bundler SfM extrae puntos característicos (*keypoints*) de un conjunto de imágenes, usando el algoritmo de *Scale-Invariant Feature Transform* (SIFT) [5]. En la Figura 3-7 se muestran los puntos característicos encontrados en dos de imágenes utilizadas en el proceso de reconstrucción de una rama.



(a) *Keypoints* encontrados por SIFT en la primera imagen. (b) *Keypoints* encontrados por SIFT en la segunda imagen.

Figura 3-7: *Keypoints* encontrados utilizando el algoritmo de SIFT.

Luego de encontrar los puntos característicos en todas las imágenes se procede a emparejar estos puntos. La Figura 3-8 ilustra el emparejamiento realizado a los puntos característicos de la Figura 3-7.



Figura 3-8: Emparejamiento de los *keypoints*.

Adicionalmente a los procesos de reconstrucción 3D descritos en la Sección 2.4, la biblioteca de Bundler SfM realiza un ajuste de Bundler con el objetivo de minimizar el error de re-proyección entre las vistas. Dicho ajuste está dado por encontrar \mathbf{P}^i y \mathbf{X}_j , que minimicen

$$\sum_{i,j} d(\mathbf{P}^i \mathbf{X}_j, \mathbf{x}_j^i)^2. \quad (3-2)$$

El ajuste de Bundler, modifica los parámetros intrínsecos de la cámara, distancia focal y distorsión, con el objetivo de minimizar el error en la re-proyección. Bundler SfM produce nubes de puntos dispersas como la mostrada en la Figura 3-9.

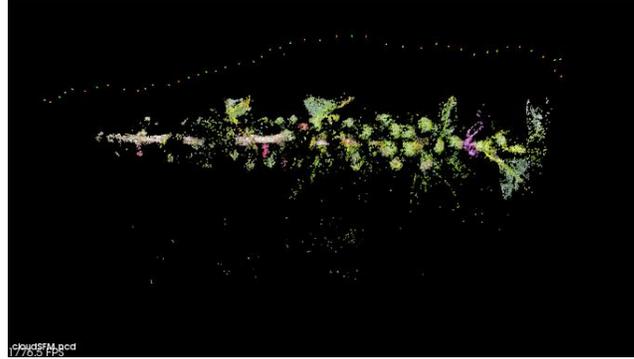


Figura 3-9: Nube de puntos generada por SfM.

Para obtener las nubes de puntos densas, se usa PMVS, el cual es un algoritmo para estéreo, que toma un conjunto de imágenes y los parámetros extrínsecos de cámara para reconstruir una estructura 3D de un objeto o una escena visible en las imágenes. Este algoritmo reconstruye únicamente estructuras rígidas. PMVS genera un conjunto de puntos orientados en lugar de un modelo poligonal, donde tanto las coordenadas 3D como la superficie normal se calcula por cada punto orientado. El modelo de parche es ilustrado en la Figura 3-10, donde un parche p es un rectángulo con centro $c(p)$ y vector normal unitario ($n(p)$) orientado hacia las cámaras que observan el rectángulo. PMVS asocia con p una imagen referencia $R(p)$, la cual se elige de tal forma que su plano es casi paralelo a p pero con distorsión. $R(p)$ determina la orientación y extensión del rectángulo p en el plano ortogonal a $n(p)$, lo que permite que la proyección de uno de los bordes de p sobre $R(p)$ sea paralela a la fila de la imagen $R(p)$. Adicionalmente, dos conjuntos de imágenes se adjuntan al parche p . El conjunto de imágenes $S(p)$, donde p debe ser visible, aunque en la práctica esto no es posible debido al movimiento en la escena, el desenfoque, los cambios de iluminación y la oclusión. Por otro lado, $T(p)$ es el conjunto de imágenes donde p se encuentra realmente en el parche p . $R(p)$ es un elemento de $T(p)$.

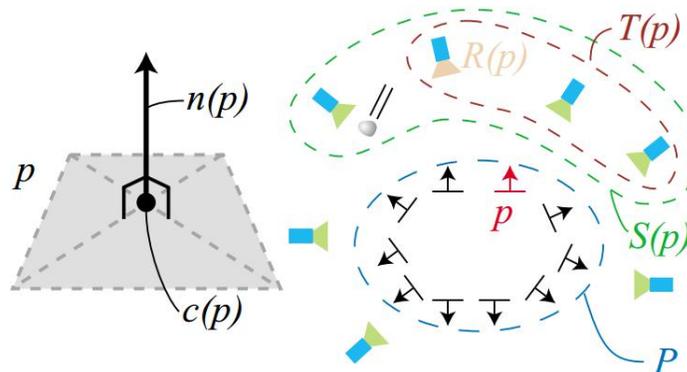


Figura 3-10: Definición de un parche (patch) y de las imágenes asociadas con el parche. Tomado de [7].

El resultado de aplicar PMVS se puede observar en la Figura 3-11.



Figura 3-11: Nube de puntos generada por PMVS

3.3. Filtrado

El objetivo de esta fase es eliminar datos que no pertenezcan a la rama, como valores atípicos, y separar las estructuras vegetativas presentes en la rama del fondo de la escena. Para esto, en primer lugar se realiza un filtrado de *outliers* que elimina valores atípicos de la nube de puntos, luego mediante el uso de un filtro pasa-banda se remueve el fondo de la escena.

3.3.1. Filtrado estadístico de outliers

Cuando se realiza la reconstrucción 3D usando SfM y PMVS se generan puntos atípicos, los cuales son errores a la hora de triangular los puntos de las imágenes. Para eliminar estos datos se realiza un análisis estadístico en el vecindario de cada punto, y posteriormente se eliminan aquellos puntos cuya distancia sea mayor a la desviación estándar de la distancia media en el vecindario. La Figura 3-12, ilustra el proceso para el filtrado de outliers sobre un conjunto de puntos dentro de un vecindario.

El Algoritmo 1, muestra el proceso para el filtrado estadístico de *outliers* sobre una nube de puntos.

Algoritmo 1 Algoritmo para el filtrado estadístico de *outliers*.**Entrada:** Nube de puntos sin filtrar. N = Número de vecinos.**Salida:** Nube de puntos Filtrada.**for** q **to** Size Input Points Cloud **do**Encontrar los N vecinos mas cercanos a q .

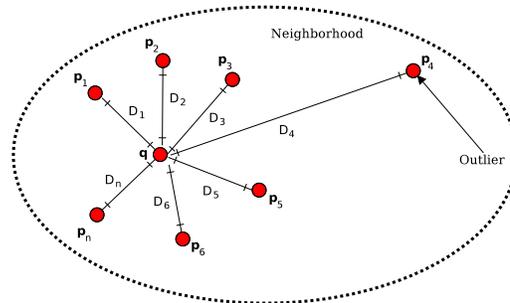
$$D(q, p_i) = \sqrt{(q_x - p_{xi})^2 + (q_y - p_{yi})^2 + (q_z - p_{zi})^2} \quad (3-3)$$

$$D_m = \frac{1}{N} \sum_{i=1}^N D(q, p_i) \quad (3-4)$$

$$\sigma_m = \sqrt{\frac{1}{N} \sum_{i=1}^N (D(q, p_i) - D_m)^2} \quad (3-5)$$

for all p_i **do****if** $D(q, p_i) > \sigma_m$ **then** $p_i \rightarrow$ Remover Punto (Outlier).**else** $p_i \rightarrow$ Guardar Punto.**end if****end for****end for**

$D(q, p_i)$ es la distancia euclidiana entre cada p_i y el punto q ; D_m es la distancia media entre puntos en el vecindario de tamaño N , y σ_m , la desviación estándar de la distancia entre cada p_i con el punto de consulta q y la distancia media D_m .

**Figura 3-12:** Filtrado Estadístico de *Outliers* sobre un vecindario.

La Figura 3-13a muestra una nube de puntos luego del proceso de reconstrucción 3D y en la Figura 3-13b se puede observar el resultado de la fase de filtrado de *outliers* sobre esta rama. La mayor cantidad de los valores atípicos se presenta en el fondo de la escena. Cada nube de puntos densa tiene un promedio de 300.000 puntos. El vecindario sobre el cual se procesó cada punto, con el objetivo de determinar si corresponde a un *outlier* o no, corresponde al 0.1 % del tamaño total de la nube de puntos.

3.3.2. Remoción de Fondo

Para remover el fondo de la escena se utilizó un filtro pasa-banda, donde se eliminan los puntos que estén por fuera de un rango específico, en la dimensión (dirección) deseada, ver Algoritmo 2. El filtro implementado trabajó sobre el eje z , debido a que el fondo se encontraba en esta dirección. Para remover un punto p , se evaluó si la componente en z del punto p , se encuentra dentro del rango Z_{min} y Z_{max} , definidos a prior. Es decir, si $Z_{min} < p_z < Z_{max}$. En la Figura 3-13c se ilustra como en esta fase se remueve el fondo de la escena dejando únicamente, en la nube de puntos, las estructuras de la rama.

Algoritmo 2 Algoritmo para la remoción de fondo utilizando un filtro pasa-banda

Entrada: Nube de puntos sin filtrar.

Salida: Nube de puntos Filtrada.

```

for  $q_i$  to Size Input Points Cloud do

  if ( $q_{iz} \geq Z_{min}$ ) and ( $q_{iz} \leq Z_{max}$ ) then
     $q_i \rightarrow$  Guardar Punto.
  else
     $q_i \rightarrow$  Remover Punto.
  end if
end for

```



(a) Nube de puntos de Entrada.

(b) Filtrado de Outliers.



(c) Remoción de Fondo.

Figura 3-13: Fase de Filtrado.

Al finalizar la etapa de filtrado se logró eliminar los valores atípicos de la reconstrucción y

el fondo de la escena. Adicionalmente, por cada nube de puntos se obtuvo una reducción del 50 % del tamaño original, lo cual redujo el tiempo computacional en los procesos posteriores, debido a que se usan menos datos, los cuales contienen la información más relevante de la escena.

3.4. Sub-muestreo

Las nubes de puntos filtradas, fueron sub-muestreadas, con el objetivo de reducir el tiempo de cómputo en los procesos de extracción de características 2D y 3D. El método implementado para este objetivo fue el filtrado por voxelización. La Figura 3-14 ilustra dicho proceso de sub-muestreo, realizado a un grano de café. En primer lugar se agrupan los puntos de la estructura en vóxeles, luego se encuentran los centroides en cada vóxel, los cuales son la nueva información de la nube de puntos sub-muestreada. El tamaño de los vóxeles (x, y, z) se selecciona de manera automática y depende del volumen de la nube de puntos, aquí se utilizaron vóxeles con un tamaño del 20 % del volumen total de la nube de puntos. Al final del proceso se genera una nueva nube de puntos en la cual se tiene un número menor de puntos, pero se mantienen las características geométricas de las ramas. En la Figura 3-15, se puede observar la nube de puntos de entrada (Figura 3-15a) y el resultado de sub-muestrear una rama de café (Figura 3-15b). La reducción del número puntos en cada rama fue de aproximadamente el 50 % del tamaño original.

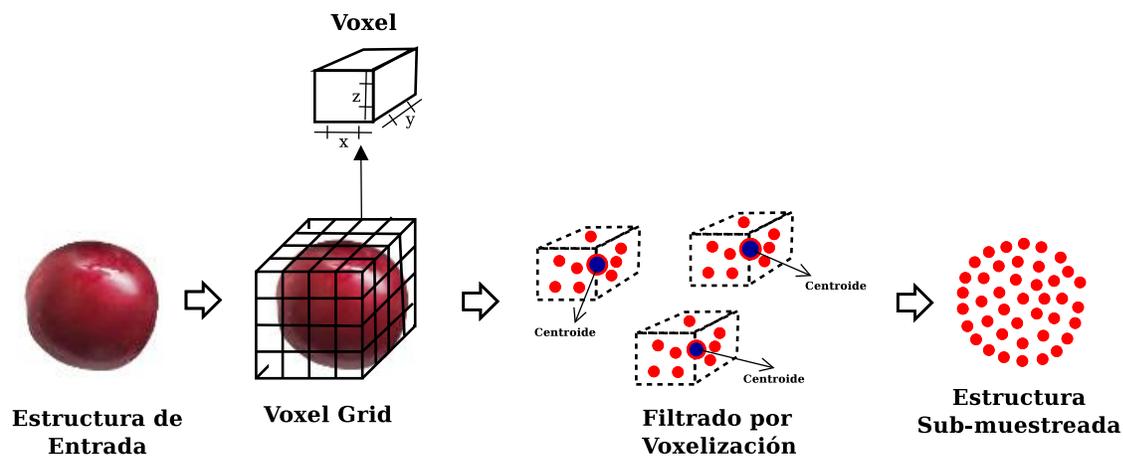
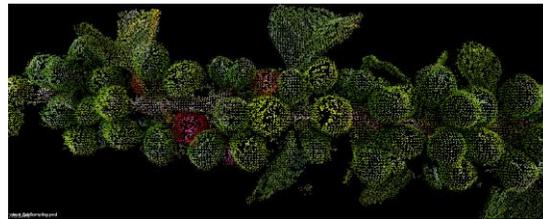


Figura 3-14: Sub-muestreo utilizando filtrado por Voxelización



(a) Nube de puntos de Entrada.



(b) Nube de puntos Submuestreada.

Figura 3-15: Fase de sub-muestreo.

Los resultados de la fase de filtrado son mostrados en las Figuras 3-16, 3-17, 3-18 y 3-19. En la parte de derecha se puede observar el resultado de la reconstrucción 3D, en el centro la rama filtrada por *outliers* y en la izquierda la rama después de haber removido el fondo.

(a) Nube de puntos de entrada. (b) Filtrado de *Outliers*. (c) Remoción de fondo.**Figura 3-16:** Resultados del proceso de filtrado de la rama No. 1(a) Nube de puntos de entrada. (b) Filtrado de *Outliers*. (c) Remoción de fondo.**Figura 3-17:** Resultados del proceso de filtrado de la rama No. 2



(a) Nube de puntos de entrada. (b) Filtrado de *Outliers*. (c) Remoción de fondo.

Figura 3-18: Resultados del proceso de filtrado de la rama No. 3



(a) Nube de puntos de entrada. (b) Filtrado de *Outliers*. (c) Remoción de fondo.

Figura 3-19: Resultados del proceso de filtrado de la rama No. 4

Por otro lado, los resultados de sub-muestrear las nubes de puntos son mostrados en las Figuras 3-20, 3-21, 3-22 y 3-23, donde se comparan las nubes de puntos resultados de la reconstrucción y las nubes de puntos sub-muestreadas. En la Tabla 3-2 se puede observar que después de realizar la fase de sub-muestreo la información se reduce en un 70 %.



(a) Nube de puntos de entrada. (b) Nube de puntos sub-muestreada.

Figura 3-20: Resultados del proceso de sub-muestreo en la rama No. 1.



(a) Nube de puntos de entrada.

(b) Nube de puntos sub-muestreada.

Figura 3-21: Resultados del proceso de sub-muestreo en la rama No. 2.

(a) Nube de puntos de entrada.

(b) Nube de puntos sub-muestreada.

Figura 3-22: Resultados del proceso de sub-muestreo en la rama No. 3.

(a) Nube de puntos de entrada.

(b) Nube de puntos sub-muestreada.

Figura 3-23: Resultados del proceso de sub-muestreo en la rama No. 4.

En la Tabla **3-2** se muestra la información del tamaño de las nubes de puntos por ramas en cada fase del proceso de filtrado. Se puede observar que la reducción en tamaño de la nube de puntos en la fase de filtrado de *outliers* es del 23% y al remover el fondo es del 31%.

Tabla 3-2: Tamaño de las nubes de puntos en cada fase del proceso de filtrado y sub-muestreo

No. de rama	Nube de puntos de entrada	Filtrado <i>Outliers</i>	Remoción de fondo	Sub-muestreo
1	225563	162427	118267	77131
2	299220	237929	233125	71015
3	269533	218206	201053	79545
4	170400	123105	116648	53341
Total	964716	741667	669093	281032

3.5. Extracción de características

3.5.1. Características 3D

Las características 3D que se extrajeron para cada estructura se basan en las curvaturas presentes en sub-regiones de cada nube de puntos sub-muestreada. En total 6 características fueron halladas: cuatro curvaturas: máxima, mínima, gaussiana y media; el índice de forma y la curvatura. Estas características fueron extraídas usando las normales, calculadas sobre regiones de cada nube de puntos. Cada región fue encontrada usando la búsqueda del vecino más cercano en una estructura de datos particionada con kd-Trees. Una estructura de datos kd-Tree organiza un número de puntos en un espacio de k dimensiones, para este trabajo se utilizaron arboles tridimensionales, $k = 3$. La Figura 3-24 ilustra el proceso para encontrar la características 3D sobre una región, en este caso un grano de café.

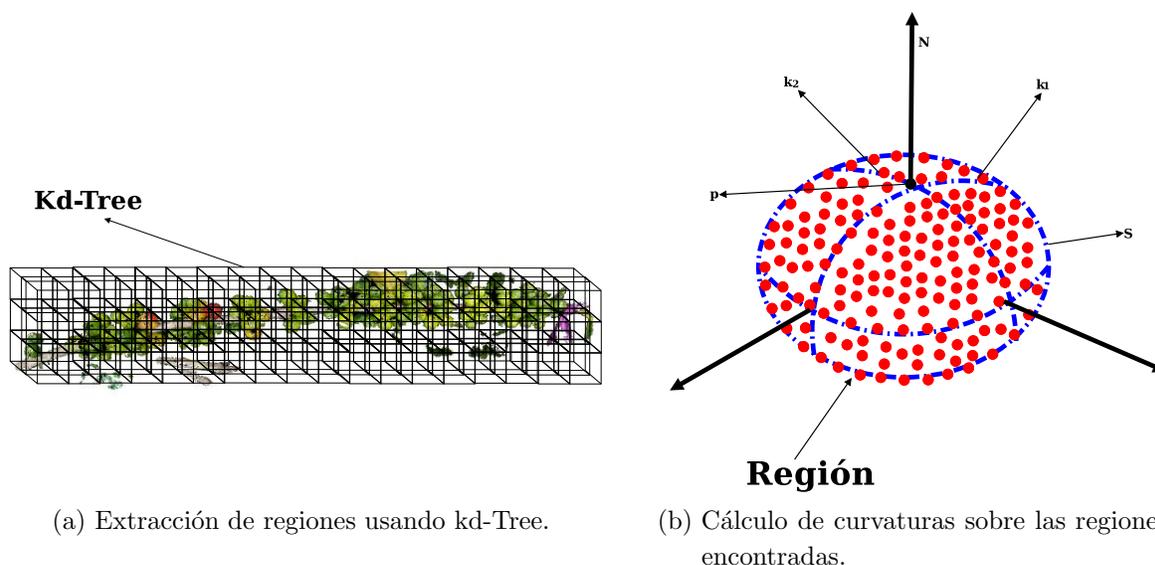


Figura 3-24: Procesos para la extracción de regiones y cálculo de curvaturas.

Para calcular la normal se realizó una descomposición en valores singulares en cada región encontrada. El vector propio que corresponde al valor propio más pequeño se aproxima a

la normal \vec{N} de la superficie S sobre el punto p , mientras que las curvaturas máxima k_1 y mínima k_2 , de la superficie S , se calculan a partir de los valores propios de la siguiente manera:

$$k_1 = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}, \quad (3-6)$$

$$k_2 = \frac{\lambda_2}{\lambda_0 + \lambda_1 + \lambda_2}, \quad (3-7)$$

dónde $\lambda_0 < \lambda_1 < \lambda_2$.

Por su parte las demás características 3D fueron calculadas utilizando las siguientes ecuaciones:

- **Curvatura Gaussiana**

$$K = k_1 k_2. \quad (3-8)$$

- **Curvatura Media**

$$H = \frac{1}{2}(k_1 + k_2). \quad (3-9)$$

- **Indice de Forma**

$$S_i = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left(\frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)} \right). \quad (3-10)$$

- **Curvividad**

$$C = \sqrt{\frac{k_1^2(p) + k_2^2(p)}{2}}. \quad (3-11)$$

3.5.2. Características 2D

Las características 2D se extraen en el momento que se realiza el sub-muestreo de la nube de puntos. A cada uno de los puntos dentro de un vóxel se le calcula la mediana de los diferentes espacios de color, RGB, Lab, Luv, YCbCr y HSV (ver (Figura 3-25)). Luego, a cada centroide se le asocia un conjunto de características que está compuesto por las componentes de color:

R, B, L, a, v, Cb, Cr, H y S. Del espacio de color RGB, se utilizaron las componentes R y G. La componente B no se utilizó debido a que en la experimentación se evidenció que al usar este componente de color se reducía el desempeño del sistema, lo anterior se presentaba por las altas variaciones de intensidad en este espacio de color, debido a las diferentes condiciones de luminosidad en que se capturó la información. Por su parte del espacio Lab, las tres componentes fueron utilizadas. De los espacios restantes no se utilizó el componente de luminosidad, debido a que esta componente ya está presente en el componente L del espacio Lab.

En las imágenes capturadas se presentan diferentes intensidades de color para un mismo punto debido a que se adquirieron bajo condiciones no controladas de luminosidad. Por esta razón, al igual que para la generación de la nube de puntos *GroundTruth*, se escogió trabajar con la mediana en vez de la media o la desviación estándar para no generar datos nuevos en las características asociadas a cada punto.

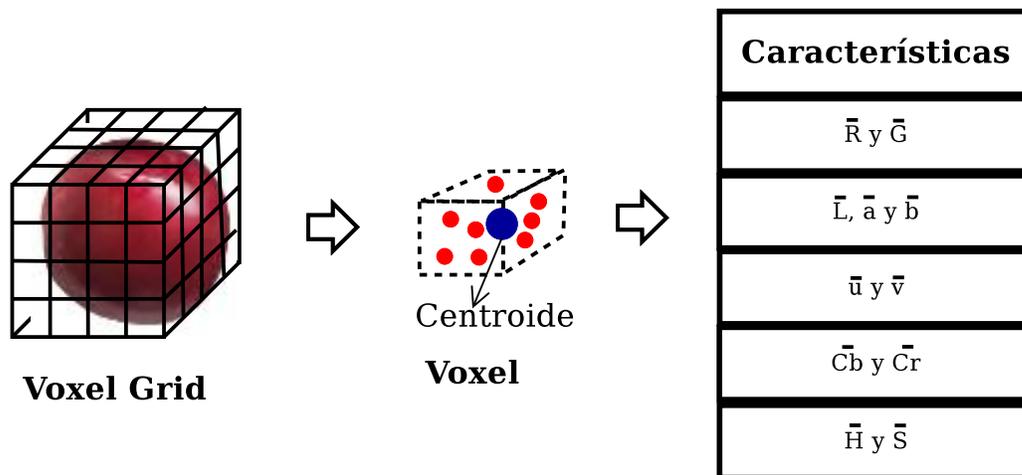


Figura 3-25: Extracción de Características 2D.

3.6. Clasificación

Para la clasificación de estructuras vegetativas se usó una máquina de vectores de soporte, empleando la biblioteca LIBSVM [58]. La SVM fue entrenada con las 17 características, 11 correspondientes a espacios de color: R, G, L, a, b, u, v, Cb, Cr, H y S; 4 de curvatura: máxima, mínima, gaussiana, y media; el índice de forma y la curvósidad. Se utilizó un kernel Gaussiano, ecuación 2-40. Para el entrenamiento del sistema se utilizaron 8 de las 12 ramas, las restantes fueron utilizadas para evaluar el desempeño del método propuesto. En total seis clases semánticas fueron clasificadas: tallo, hoja, flor, fruto inmaduro, fruto semi-maduro y fruto maduro. La salida del clasificador es una nube de puntos, donde cada punto es asignado a una de las seis clases.

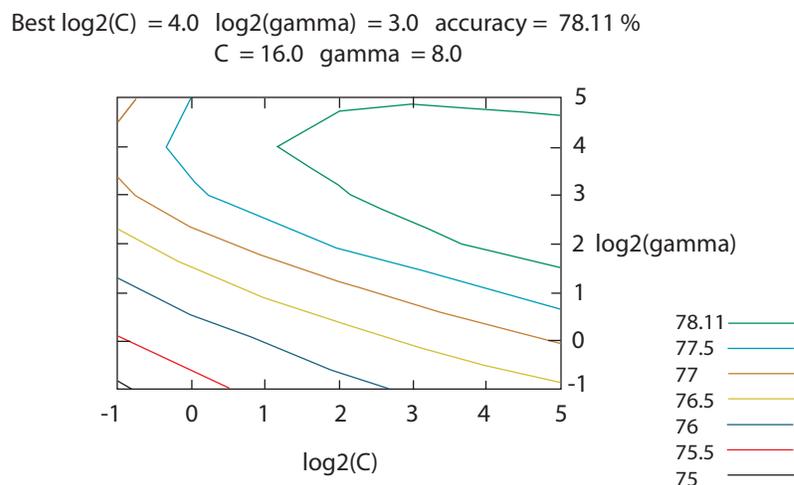


Figura 3-26: Proceso de validación cruzada usando un tamaño de región = 2.5cm.

Para sintonizar los parámetros C de la ecuación 2-39 y γ de la Ecuación 2-40, se realizó un proceso de validación cruzada. La Figura 3-26 muestra los resultados obtenidos después del proceso de validación cruzada, donde se variaron los valores de C y de γ en potencias de 2, y se buscó los valores para C y γ en los cuales se obtuvo el mejor desempeño. En este caso, la región usada para la extracción de características 3D fue 2.5 cm. Los valores óptimos para C y γ son 16 y 8 respectivamente.

Se extendió el modelo binario de SVM descrito en la ecuaciones 2-38 y 2-39, a una clasificación de varias clases usando el método conocido como uno contra el resto. Se construyeron k clasificadores binarios $f_i (i = 1, \dots, k)$ que separan las clases i del resto de clases $k - 1$. La regla de decisión para un problema con k -clases es:

$$y(x) = \arg \max_{1 \leq i \leq k} f_i(x) \quad (3-12)$$

dónde f_i es la función de decisión descrita en la Ecuación 2-41.

4 Resultados

En este trabajo se adquirió información de 12 ramas de café. Por cada rama se capturaron vídeos y se realizaron los procesos de: reconstrucción 3D, filtrado, sub-muestreo, extracción de características y generación de la información *Groundtruth*, expuestos en la sección anterior. Para la fase de entrenamiento de la SVM se utilizaron 8 ramas, las 4 ramas restantes fueron usadas para validar el desempeño del clasificador. En la Figura 4-1, se ilustran las 4 ramas utilizadas para la validación del método propuesto, estas ramas fueron seleccionadas de tal forma que se garantice datos para las seis clases o estructuras vegetativas de interés. La Tabla 4-1 muestra el porcentaje de datos utilizados por clase para las fases de entrenamiento y la validación, se evidencia que el mayor porcentaje de datos se encuentra en la clase fruto inmaduro, con un 64,3 % y 51.76 % en las fases de entrenamiento y validación. Lo anterior se debe a que en las ramas de café hay mayor cantidad de información de frutos inmaduros que de otra clase.

Tabla 4-1: Porcentaje de datos utilizado por clase en las fases de entrenamiento y validación.

Clase	Entrenamiento	Validación
Fruto maduro	4.52 %	2.99 %
Tallo	14.03 %	14.36 %
Hoja	12.59 %	27.88 %
Flor	2.58 %	1.42 %
Fruto inmaduro	64.30 %	51.76 %
Fruto semi-maduro	1.98 %	1.59 %

Para tratar el des-balance entre clases, se implementó una extensión de las SVM conocida como: Weighted SVM o (WSVM) [58, 59, 60], donde se aplica una penalización diferente por cada clase, aplicando mayor penalidad a las clases con mayor número de muestras. En WSVM el modelo estándar de SVM presentado en las ecuaciones 2-38 y 2-39 se convierten en:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \mu_i \xi_i, \quad (4-1)$$

Sujeto a:

$$y_i(\langle \mathbf{W}, \phi(x_i) \rangle + b) + \xi_i \geq 1, \quad i = 1, 2, \dots, l, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, l.$$

El modelo dual estaría dado por:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j k(x_i, x_j) - \sum_{i=1}^l \alpha_i \quad (4-2)$$

Sujeto a:

$$\sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq \mu_i C, \quad i = 1, 2, \dots, l,$$

dónde $\mu_i (i = 1, \dots, l)$ indica el peso de cada clase x_i .

Aunque existe un des-balance en lo datos usados por clase, el uso WSVM permitió que el desempeño del sistema no se afecte por el des-balance en el tamaño de las clases. Por ejemplo, la clase fruto maduro tiene un porcentaje de 4.52 % de datos para el entrenamiento. Sin embargo, al mirar la Tabla 4-2 se puede observar que uno de los mejores desempeños del sistema fue obtenido para la clase fruto maduro con un F1-Score entre 0.7 y 0.83. Lo anterior sugiere que el sistema propuesto no depende directamente del número de datos usados por clase para el entrenamiento, sino que el desempeño depende del tamaño de región sobre el cual se extraen las características.

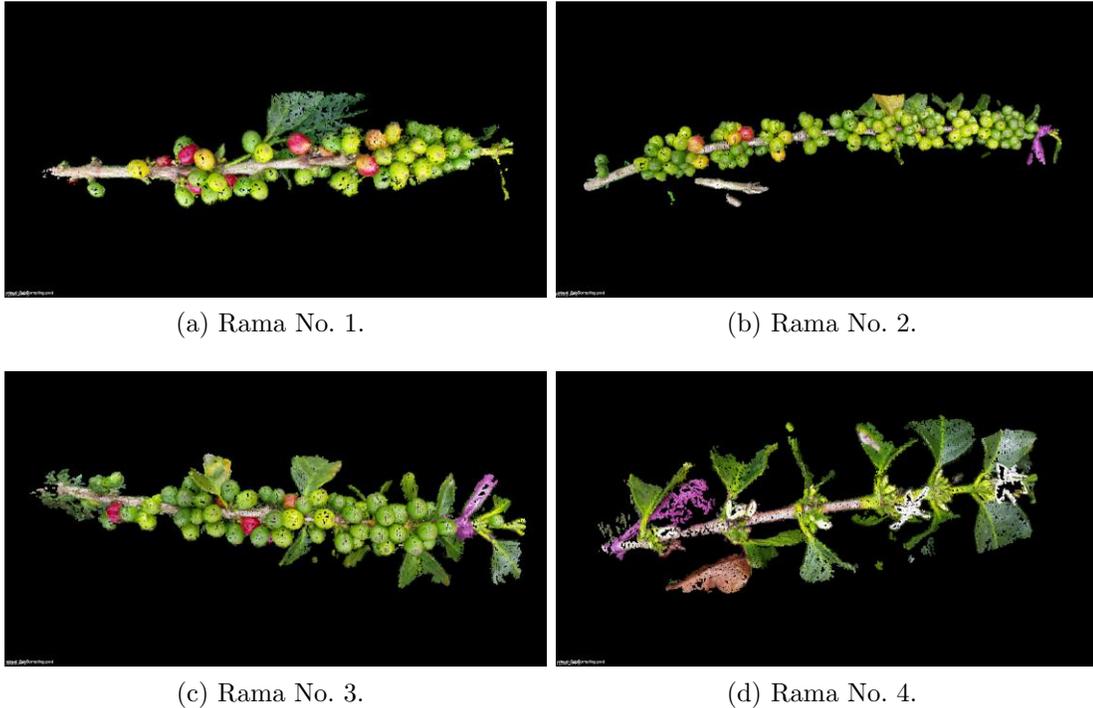


Figura 4-1: Ramas utilizadas para la validación del sistema.

4.1. Índices de evaluación

Para evaluar el desempeño del sistema propuesto, se utilizó la matriz de confusión, y las métricas de exactitud total, precisión, recall o especificidad y F1-Score, ecuaciones 4-3, 4-4, 4-5 y 4-6. La exactitud relaciona los datos correctamente clasificados y el total de datos evaluados. La precisión mide la relación que existe entre la cantidad de datos correctamente clasificados en una clase y el número de datos clasificados como esa clase. El recall mide la relación que existe entre los datos correctamente clasificados y el número real de datos pertenecientes a esa clase. Finalmente, con el objetivo de no sesgar la evaluación del sistema se calculó el índice F1-Score, el cual entrega una ponderación entre la precisión y recall. Las anteriores métricas son definidas a continuación.

$$\text{Exactitud} = \frac{\sum_i^n cm_{ii}}{\sum_i^n \sum_j^n cm_{ij}}. \quad (4-3)$$

$$\text{Precision} = \frac{cm_{ii}}{\sum_j^n cm_{ji}}. \quad (4-4)$$

$$\text{Recall} = \frac{cm_{ii}}{\sum_j^n cm_{ij}}. \quad (4-5)$$

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (4-6)$$

Donde, cm es la matriz de confusión construida con los resultados de la clasificación y los datos *Groundtruth*, y n es el número de clases.

4.2. Clasificación de las estructuras vegetativas

En el Sección3.6 de este documento, se explicó el proceso para la extracción de características. Para el caso de las características 3D fue necesario determinar el tamaño de las regiones sobre las cuales se extraen. La Tabla 4-2 muestra el índice F1-Score por clase obtenido por el sistema, al variar el tamaño de la región. La variación del tamaño de la región se realizó en un rango de 0.5 a 5 centímetros (cm).

Tabla 4-2: F1 - Score por clase vs. tamaño de región

Clase	Tamaño de la Región (cm)							
	0.5	1.0	1.5	2.0	2.5	3.0	4.0	5.0
Fruto maduro	0.70	0.84	0.85	0.85	0.85	0.83	0.81	0.79
Tallo	0.33	0.35	0.72	0.73	0.73	0.73	0.73	0.73
Hoja	0.00	0.00	0.81	0.22	0.85	0.85	0.85	0.85
Flor	0.36	0.63	0.73	0.74	0.74	0.72	0.64	0.59
Fruto inmaduro	0.04	0.16	0.75	0.59	0.77	0.77	0.77	0.77
Fruto semi-maduro	0.07	0.17	0.19	0.21	0.19	0.17	0.16	0.15

A continuación se muestran las matrices de confusión resultantes con cada tamaño de región utilizado.

Clase Real	F. Maduro	1170	690	0	26	0	26
	Tallo	3	9953	0	2457	15	341
	Hoja	0	19317	0	3262	719	8
	Flor	0	73	0	1854	0	4
	F. Inmaduro	12	16553	0	644	409	2996
	F. Semi-maduro	248	246	0	24	0	138
		F. Maduro	Tallo	Hoja	Flor	F. Inmaduro	F. Semi-maduro
		Clase Estimada					

Figura 4-2: Matriz de confusión resultante al utilizar un tamaño de región de 0.5cm

Clase Real	F. Maduro	1685	166	0	7	1	53
	Tallo	3	10824	0	1600	310	32
	Hoja	0	21478	0	79	1749	0
	Flor	1	86	0	1728	104	12
	F. Inmaduro	65	17087	0	170	1925	1367
	F. Semi-maduro	337	101	0	4	22	192
		F. Maduro	Tallo	Hoja	Flor	F. Inmaduro	F. Semi-maduro
		Clase Estimada					

Figura 4-3: Matriz de confusión resultante al utilizar un tamaño de región de 1.0cm

Clase Real	F. Maduro	1708	49	0	4	126	25
	Tallo	1	8205	372	926	3221	44
	Hoja	0	890	16570	32	5813	1
	Flor	0	40	10	1732	137	12
	F. Inmaduro	73	736	473	121	17927	1284
	F. Semi-maduro	337	28	0	7	75	209
		F. Maduro	Tallo	Hoja	Flor	F. Inmaduro	F. Semi-maduro
		Clase Estimada					

Figura 4-4: Matriz de confusión resultante al utilizar un tamaño de región de 1.5cm

Clase Real	F. Maduro	1732	33	0	2	112	33
	Tallo	4	8443	185	810	3272	55
	Hoja	0	831	2920	47	19508	0
	Flor	0	43	3	1696	179	10
	F. Inmaduro	83	889	92	117	18201	1232
	F. Semi-maduro	354	21	0	7	43	231
		F. Maduro	Tallo	Hoja	Flor	F. Inmaduro	F. Semi-maduro
		Clase Estimada					

Figura 4-5: Matriz de confusión resultante al utilizar un tamaño de región de 2.0cm

Clase Real	F. Maduro	1710	29	0	4	117	52
	Tallo	3	8481	364	837	3021	63
	Hoja	0	850	17894	40	4522	0
	Flor	0	51	6	1748	120	6
	F. Inmaduro	54	918	541	155	17704	1242
	F. Semi-maduro	356	45	0	10	36	209
		F. Maduro	Tallo	Hoja	Flor	F. Inmaduro	F. Semi-maduro
		Clase Estimada					

Figura 4-6: Matriz de confusión resultante al utilizar un tamaño de región de 2.5cm

Clase Real	F. Maduro	1674	34	0	22	133	49
	Tallo	2	8389	372	886	3063	57
	Hoja	0	751	17920	42	4593	0
	Flor	0	57	3	1718	134	19
	F. Inmaduro	69	789	410	148	17904	1294
	F. Semi-maduro	355	42	0	13	48	198
		F. Maduro	Tallo	Hoja	Flor	F. Inmaduro	F. Semi-maduro
		Clase Estimada					

Figura 4-7: Matriz de confusión resultante al utilizar un tamaño de región de 3.0cm

Clase Real	F. Maduro	1606	33	0	17	219	37
	Tallo	15	8158	292	1293	2946	65
	Hoja	0	1155	17917	94	4140	0
	Flor	0	77	1	1621	227	5
	F. Inmaduro	35	1046	396	116	17675	1346
	F. Semi-maduro	388	26	1	12	51	178
		F. Maduro	Tallo	Hoja	Flor	F. Inmaduro	F. Semi-maduro
		Clase Estimada					

Figura 4-8: Matriz de confusión resultante al utilizar un tamaño de región de 4.0cm

Clase Real	F. Maduro	1573	29	0	12	275	23
	Tallo	4	7993	344	1527	2881	20
	Hoja	0	846	17762	88	4610	0
	Flor	0	84	2	1529	306	10
	F. Inmaduro	34	868	462	109	17842	1299
	F. Semi-maduro	372	39	0	1	83	161
		F. Maduro	Tallo	Hoja	Flor	F. Inmaduro	F. Semi-maduro
		Clase Estimada					

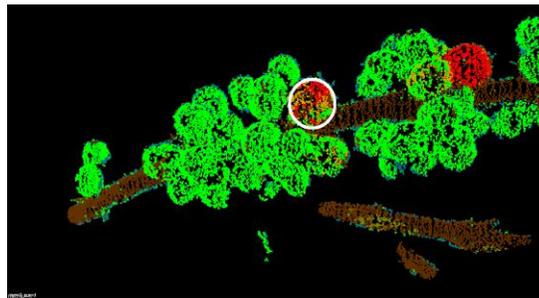
Figura 4-9: Matriz de confusión resultante al utilizar un tamaño de región de 5.0cm

Al examinar el desempeño que obtuvo el sistema por clase se observa lo siguiente:

- Para las clases fruto inmaduro y maduro, el mejor desempeño se obtiene al utilizar un tamaño de región de 2.5 centímetros, donde se obtiene un F1-Score de 0.85 para la clase fruto maduro y 0.77 para la clase fruto inmaduro. Los resultados anteriores son razonables debido a que el rango en el que puede variar el tamaño de un grano de café en estos estados de maduración es de 1 a 2.5 centímetros. Después de este tamaño de región, el desempeño del sistema empieza a disminuir debido a que las regiones que encuentra el sistema no se ajustan a la geometría de los frutos.
- En la clase tallo, se observa que el sistema obtiene un F1-Score de 0.72 a partir de un tamaño de 1.5 centímetros, de ahí en adelante el desempeño del sistema se mantiene. Lo anterior se presenta porque la longitud de un tallo, entre cada glomerulo, oscila entre 1.5 y 5 centímetros.
- La clase hoja por su parte, tiene un bajo desempeño para las curvaturas en regiones con tamaños de 0.5, 1 y 2 centímetros. Para los tamaños de región 0.5 y 1 centímetros, se obtiene un F1-Score de 0.0 debido a que el sistema confunde parte de los frutos inmaduros con las hojas, ver figuras 4-2 y 4-3. Lo anterior se presenta debido a que la forma de las regiones de los frutos inmaduros que se encuentran se asimilan a la forma de las hojas. Caso similar ocurre para el caso de 2 cm, donde se obtiene un F1-Score de 0.22, pero en esta ocasión las regiones de las hojas son las que se asimilan a las de los frutos inmaduros (Figura 4-5).

- Las flores presentes en las ramas de café procesadas se encuentran en su etapa de flor abierta, donde cada flor puede medir entre 2.0 y 3 centímetros de radio. Se puede comprobar que para la clase flores el mejor desempeño se obtiene en un rango de 2 a 2.5 centímetros con un F1-Score de 0.74. También se puede evidenciar que a medida que crece el tamaño de la región, el desempeño para la clase flor disminuye debido a que con regiones más grandes el sistema no se ajusta a la geometría de los flores.
- Al igual que con la clase hoja, para la clase frutos inmaduros se obtienen los desempeños más bajos en los tamaños de región de 0.5, 1.0 y 2.0 centímetros con F1-Score de 0.04, 0.16, 0.59 respectivamente, debido a que el sistema confunde estas dos clases. Los frutos inmaduros pueden medir de 1 a 2.5 centímetros, dependiendo de su crecimiento y desarrollo. Para las ramas utilizadas en los experimentos, este radio oscilaba entre 1.5 y 2.5 centímetros. La Tabla 4-2 muestra que el mejor desempeño para la clase fruto inmaduro se obtiene a partir de un radio de 1.5 centímetros, donde se obtuvo un F1-Score de 0.77.
- Los frutos semi-maduros generalmente presentan zonas en las cuales las tonalidades de color se asimilan a la de los frutos maduros e inmaduros. Tal como se muestra en la Figura 4-10b, donde se encierra en un círculo blanco un fruto semi-maduro. Debido a lo anterior la clase con menor desempeño es la clase fruto semi-maduro, donde el mayor F1-Score, es de 0.21, y se obtiene al utilizar un tamaño de región de 2 centímetros. Lo anterior es razonable, debido a que el tamaño de un fruto semi-maduro debe estar entre el tamaño de un fruto inmaduro y un fruto maduro, teniendo en cuenta las curvas de crecimiento y desarrollo de los frutos 2-3. La figuras 4-2 a 4-9 muestran que los frutos semi-maduros se confunden con frutos maduros y a medida de que el tamaño de la región crece, se empiezan a confundir con los frutos inmaduros, esto se debe a que los frutos semi-maduros tienen características de color y curvaturas que se asemejan a las clases fruto maduro y fruto semi-maduro. En la Figura 4-10 se observa la clasificación del sistema en un fruto semi-maduro, el cual está encerrado en un círculo blanco, junto con una de las imágenes fuente de las cuales se obtuvo la reconstrucción 3D. Se puede evidenciar que el sistema separa el fruto de las demás estructuras presentes en la escena, esto se logra gracias a las características 3D utilizadas, debido a que las curvaturas en las diferentes estructuras como por ejemplo: tallos, hojas y flores son diferentes a las de los frutos. Al evaluar visualmente los resultados obtenidos en esta estructura, en la Figura 4-10a, se evidencia que el sistema separó el fruto pero clasificó zonas como fruto inmaduro, semi-maduro y maduro.
- Es importante analizar la causa por la cual el desempeño del sistema se mantiene para ciertas clases, aun cuando se aumenta el tamaño de las regiones con el cual se extraen las características 3D. Se puede observar que para el caso de las clases: tallo, hoja y fruto inmaduro, el desempeño del sistema converge y se mantiene cuando

se aumenta el tamaño de la región. Lo anterior no ocurre para las restantes clases: fruto maduro, fruto semi-maduro y flores, donde al aumentar el tamaño de la región el desempeño del sistema empieza a disminuir. Si se analiza la Tabla 4-1, donde se relacionó el porcentaje de datos por clase utilizados, se puede evidenciar que las clases donde el desempeño disminuye cuando se aumenta el tamaño de la región, son las clases con menos datos. Como el objetivo del sistema propuesto es clasificar estructuras vegetativas presentes en ramas de café a partir de videos capturados en campo por un dispositivo móvil, las ramas utilizadas fueron escogidas aleatoriamente de un cultivo. La naturaleza de las ramas de café utilizadas en su mayoría guardaron mayor proporción de frutos inmaduros que de otros estados. Aunque, las ramas con flores no contenían estructuras de frutos y tenían mayor presencia de estructuras como hojas y tallos que de flores. Al aumentar el tamaño de la región sobre el cual se extraen las características las clases con menor presencia de datos empiezan a tener un mal desempeño, debido a que las regiones procesadas no se acomodan a clases conocidas y el sistema empieza a confundirse en estas regiones.



(a) Anotación fruto semi-maduro



(b) Imagen fuente

Figura 4-10: Análisis del resultado de anotación fruto semi-maduro.

Las seis estructuras vegetativas presentes en las ramas de café, tienen un tamaño promedio de 2.5 centímetros. Por ejemplo, las flores presentes en las ramas procesadas, tienen un tamaño que oscila entre 1 y 3 centímetros; los frutos de café pueden medir de 1cm a 2cm, dependiendo del estado de desarrollo. Por su parte los fragmentos de tallo tienen una longitud que varía de 1.5 a 4cm entre cada glómulo. Finalmente, los fragmentos de las hojas

presentes en la escena miden de 2 a 5cm. Al analizar la Figura 4-11, donde se muestran los índices de evaluación promedio de todas las clases al variar el tamaño de la región sobre el cual se extraen las características 3D, se puede observar que el mejor desempeño se obtuvo al trabajar con un radio de 2.5 centímetros, donde se obtuvo un F1 -Score de 0.79, una precisión de 0.81 y un recall de 0.78.

Los resultados, al utilizar un tamaño de región de 2.5 centímetros, son mostrados en la Figura 4-12. En la parte izquierda se encuentran las nubes de puntos *GroundTruth*, y en la derecha las nubes de puntos anotadas por el sistema. Se puede observar que la clasificación realizada por el sistema es buena, debido a que se logran identificar visualmente las diferentes estructuras vegetativas de las ramas. Sin embargo, existen zonas que no están correctamente clasificadas, como por ejemplo: la hoja de la rama 1 (Figura 4-12(b)), los tallos cercanos a las flores en la rama 4 (Figura 4-12(h)) y los frutos semi-maduros en las ramas 1, 2 y 3 (Figura 4-12(b),(c) y (d)).

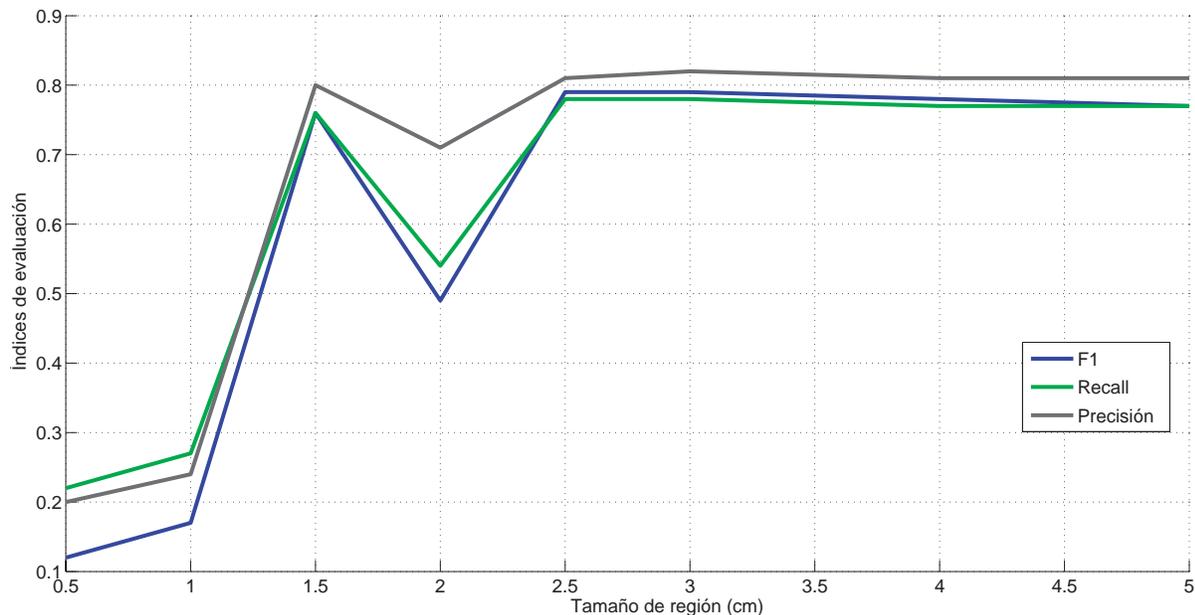
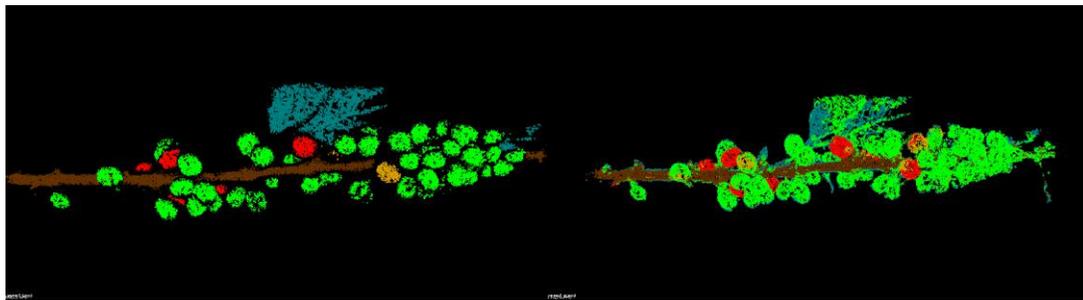
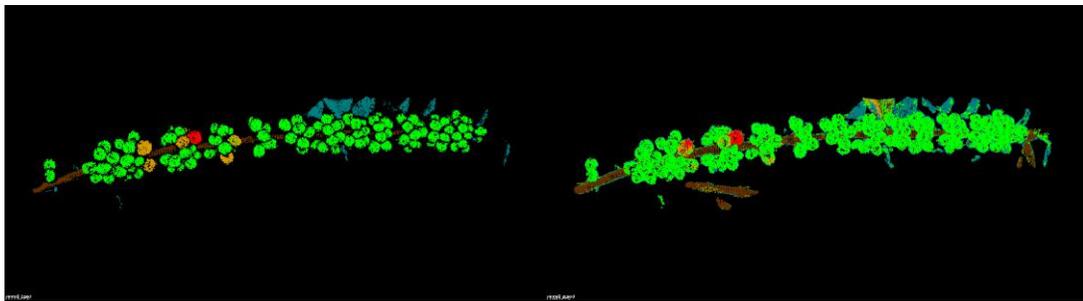


Figura 4-11: Tamaño región vs. Índices de evaluación.



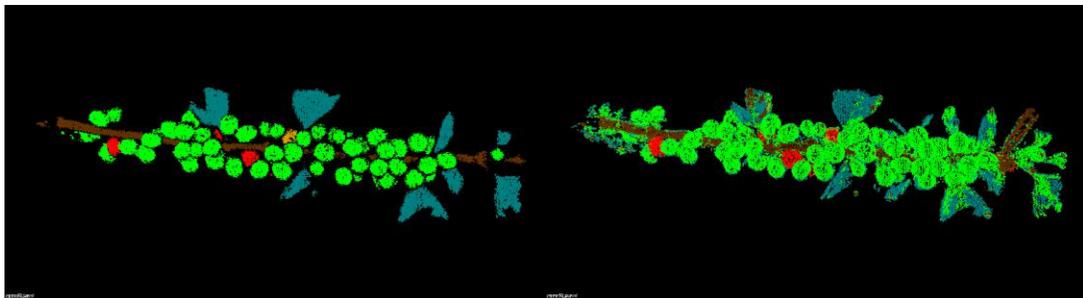
(a) Groundtruth rama No. 1.

(b) Resultados rama No. 1.



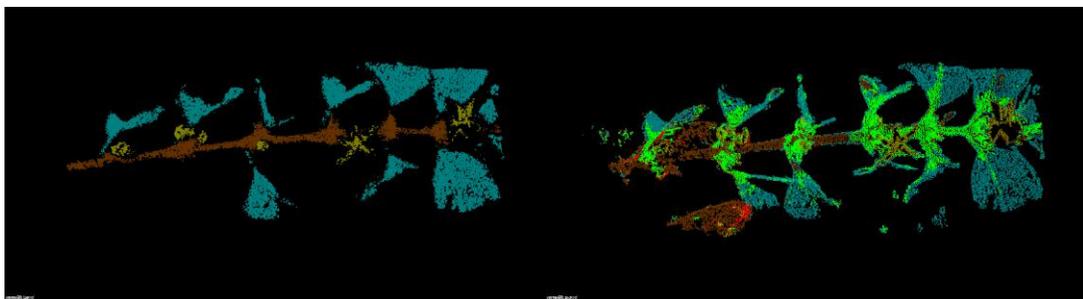
(c) Groundtruth rama No. 2.

(d) Resultados rama No. 2.



(e) Groundtruth rama No. 3.

(f) Resultados rama No. 3.



(g) Groundtruth rama No. 4.

(h) Resultados rama No. 4.

Figura 4-12: Nubes de puntos GroundTruth vs. los resultados obtenidos por el sistema propuesto.

Otro experimento que se realizó, fue unir las clases frutos inmaduros, semi-maduros y ma-

duros en una sola clase llamada frutos y evaluar nuevamente el desempeño del sistema, utilizando el mismo tamaño de región de 2.5 centímetros. Los resultados de este experimento son mostrados en la matriz de confusión de la Figura 4-13, donde se obtiene una exactitud del 94 % clasificando frutos, esto es útil para aplicaciones donde se busque realizar un conteo de los frutos, por ejemplo para la estimación de masa de café.

Clase Real	Frutos	109410	4531	1293	676
	Tallo	16240	42056	1314	4235
	Hoja	68508	4198	43606	218
	Flor	787	235	25	8608
		Frutos	Tallo	Hoja	Flor
		Clase Estimada			

Figura 4-13: Matriz de confusión uniendo las clases: fruto inmaduro, semi-maduro y maduro.

El sistema propuesto fue implementado sobre un computador personal con un procesador i5 a 3.6Ghz y una memoria RAM de 8Gb. Los tiempos promedios empleados por el sistema en cada etapa para procesar una rama son mostrados en la Tabla 4-3. Los procesos de mayor costo computacional son la reconstrucción 3D, seguido de la clasificación. El tiempo promedio requerido por el sistema para entregar una rama anotada es de 30 minutos aproximadamente. Los tiempos que se muestran en la Tabla 4-3, no incluyen el tiempo empleado para el entrenamiento del sistema el cual fue de 48 horas y 23 minutos.

Tabla 4-3: Tiempo empleado por rama en cada etapa del sistema propuesto.

Etapa	Tiempo (Seg.)
Reconstrucción 3D	941
Filtrado	131
Sub-muestreo	103
Extracción de Características	65
Clasificación	615
Total	1855

4.3. Discusión

La Tabla 4-4 muestra los resultados de algunos desarrollos y aplicaciones de visión por computador existentes, los cuales trabajan en campo y en algunos casos bajo condiciones de luminosidad no controladas como es el caso de este trabajo. Se puede evidenciar que, al utilizar únicamente características 2D, se presentan inconvenientes con las condiciones de luminosidad, oclusión y ruido [42, 34, 35]. Mientras que al usar información 3D se genera cierta independencia a estos inconvenientes. Sin embargo, en algunos casos es necesario utilizar la información 2D, debido a que los datos geométricos (3D) no son relevantes para la clasificación de las estructuras vegetativas. Por ejemplo en el caso del café, donde los frutos en sus diferentes estados de desarrollo pueden tener la misma forma geométrica pero diferentes colores.

El trabajo presentado en [61], presenta un método para contar frutos presentes en ramas de café, a partir de imágenes capturadas en condiciones de campo por un dispositivo móvil. El sistema propuesto muestra un buen desempeño, con un $R^2 = 0,82$. Sin embargo, una desventaja de este trabajo fue el no poder remover el fondo de la escena. El trabajo presentado en esta tesis trabaja con información 3D y es capaz de separar la rama del fondo de la escena, de esta forma se mitiga los problemas presentados en el trabajo desarrollado en [61].

Algunos de los trabajos listados en la Tabla 4-4, se han aplicado a diferentes cultivos como por ejemplo: tomates [35], piña [46], uvas [44], y manzanas [36]. Dichos trabajos se han aplicado en cultivos donde la disposición de las estructuras vegetativas permite una mejor separación de clases que al trabajar en cultivos de café. En las ramas de café los frutos se encuentran en racimos y sus colores son muy similares entre si, por ejemplo, los frutos semi-maduros dependiendo de su crecimiento y desarrollo son similares en tamaño y forma a los frutos maduros e inmaduros. En cuanto a color, un fruto inmaduro maneja diferentes proporciones de color. Por ejemplo se puede presentar un 50% de tonalidad verde y otro 50% de tonalidad roja. Por otra parte, en ocasiones las intensidades de color de las hojas son similares a los frutos inmaduros.

Los trabajos relacionados en la Tabla 4-4 generalmente buscan clasificar solamente los frutos, a diferencia de este trabajo donde se busca separar diferentes estructuras como: fruto en sus diferentes estados de maduración, hojas, flores y tallos presentes en cada rama. También cabe resaltar la diferencia en el proceso de adquisición de los datos en campo, en este trabajo se utilizó únicamente un dispositivo móvil y se adquirieron imágenes en movimiento con diferentes condiciones de luminosidad no controladas. Con un F1-Score del 0.79, el sistema propuesto se puede considerar adecuado para la clasificación de estructuras vegetativas presentes en ramas de café, debido a que integra información 2D y 3D, que trabaja bajo diferentes condiciones de luminosidad y ruido en ambientes exteriores, y que es accesible

para los caficultores debido a que para la captura de información únicamente se necesita de la cámara de un dispositivo móvil.

Tabla 4-4: Tabla comparativa de algunos trabajos relacionados con el método propuesto.

Referencia	Tipo de Cultivo	Características usadas	Problemas	Número de Estructuras Segmentadas	Radio promedio (cm)	Desempeño
Moonrinda et al. [46]	Piñas	SIFT	Ambigüedad SfM	(1) Frutos	10	error $\approx 29,45\%$
Patel et al. [42, 34]	Cerezas, uvas, manzanas, peras, fresas, naranjas, melocotones, etc.	Color y forma	Luminosidad	(1) Frutos	3	Eficiencia = 98%
Hunt et al. [41]	Cereal	Color infrarrojo	Costo computacional	(1) Cultivo.	No reporta	r = 0,73
Dey et al. [44]	Uvas	2D: Color y curvaturas	Ruido inducido por la clasificación independiente de cada punto	(3) Frutos, hojas y tallos	3	Eficiencia = 84,8%
Verma et al. [35]	Tomates	Forma	Oclusión	(1) Frutos	5	error $\leq 10\%$
Jay et al. [37]	Col, coliflor, girasol y remolacha	Altura e índice de vegetación	Movimiento de las estructuras por el viento	(1) Hojas	20 - 60	r = 0,94
Gongal et al. [36]	Manzanas	Color y ubicación espacial de los frutos	No reporta	(1) Frutos	2 - 3	Eficiencia = 79,8%
Ramos et al. [61]	Café	Color y Forma	Remoción de fondo	(4) Frutos	1 - 2	$R^2 = 0,82$
Método Propuesto	Café	Color, curvaturas, curvósidad y forma	Similaridad de las curvaturas y los espacios de color en algunas zonas de la rama y Flor	(6) Fruto Verde, Fruto Rojo, Fruto Pintón, Hoja, Tallo y Flor	3-4	F1-Score = 0,79%

5 Conclusiones y Trabajos Futuros.

Este trabajo presentó el desarrollo de un sistema para la clasificación de las estructuras vegetativas: flores, tallo, hojas y frutos inmaduros, semi-maduros y maduros, presentes en ramas de café, usando como fuente de información secuencias de imágenes adquiridas bajo condiciones de campo por la cámara de un dispositivo móvil.

El uso de las técnicas de SfM y PMVS permitió obtener una solución económica para la adquisición de información 3D en campo, debido a que para la reconstrucción 3D de las ramas se necesitó únicamente los vídeos adquiridos por el dispositivo móvil, en este caso un celular, a lo largo de las ramas.

Generalmente, en las imágenes capturadas en campo sobre ramas de café se presentan hojas secas, malas hierbas y ramas de otros árboles que hacen parte del fondo de la escena. Este fondo genera dificultad para la clasificación de las estructuras vegetativas de las ramas, debido a que se presentan estructuras no homogéneas que confunden los sistemas de clasificación automática. Al utilizar información 3D fue posible remover el fondo de la escena, lo que permitió clasificar únicamente las estructuras presentes en la rama.

Debido a las dificultades, con las condiciones de luminosidad y oclusión, que se presentan cuando se trabaja con imágenes adquiridas en campo, este trabajo hizo uso de características de color, curvaturas y los índices de forma y curvatura, para la clasificación de las diferentes estructuras vegetativas presentes en la escena. Aunque las imágenes adquiridas tenían diferentes condiciones de luminosidad y ruido, el sistema propuesto mostró independencia a los cambios de luminosidad y oclusión. Los resultados experimentales mostraron una exactitud global del 81 % y un F1-Score de 0.79 sobre los datos de validación.

Al analizar el desempeño del sistema por cada clase, se evidenció que el tamaño de la región sobre el que se calculan las características 3D afecta directamente el resultado de la clasificación. El sistema mostró estar relacionado con el tamaño de las estructuras. Por ejemplo, para separar los frutos de las otras estructuras se debe utilizar un tamaño de región que este entre 1.5 y 2.5 centímetros, el cual es el rango real de tamaños en el cual puede estar un grano de café. El tamaño real promedio de las estructuras vegetativas: hojas, flores, tallos, frutos inmaduros, semi-maduros y maduros es 2.5 centímetros. El mejor desempeño del sistema se obtuvo al trabajar con un radio de 2.5 centímetros, donde se obtuvo una precisión

de 0.81 y un recall de 0.78, lo que indica que el método propuesto tiene un buen desempeño clasificando las estructuras vegetativas presentes en las ramas de café. Sin embargo, se pudo evidenciar que existen estructuras que por su forma y color no permiten una buena clasificación, debido a que son muy similares a otras estructuras. Por ejemplo, existen zonas en los frutos inmaduros, que por su curvatura y color se asimilan a las hojas. Por su parte, los frutos semi-maduros, debido al porcentaje en cantidad de color rojo y verde, se asimilan en algunas ocasiones a los frutos inmaduros y en otras a los frutos maduros.

En la Figura 4-12 se pueden observar los resultados obtenidos por el sistema propuesto, donde se mitigó los problemas de oclusión de las estructuras vegetativas presentes en cultivos de café. Al utilizar información 3D fue posible trabajar con la ubicación espacial de los frutos, lo que permitió identificar frutos aun cuando estos estaban ocluidos por otras estructuras.

Para mejorar el desempeño del sistema, se propone la inclusión de diferentes características basadas en textura que complementen las utilizadas en el presente trabajo, y permitan una mejor clasificación de la clase fruto semi-maduro. Aunque el sistema propuesto extrae características sobre regiones de la nube de puntos, se propone implementar *Conditional Random File* (CRF) para reducir el ruido inducido por la clasificación independiente de cada punto. Finalmente, el sistema propuesto será implementado sobre un servidor dedicado con el objetivo de reducir los tiempos de cómputo por rama.

Bibliografía

- [1] B. R. H. M. A. P. J. J. R. A. Salazar G., M.R.; Chaves C., “Crecimiento del fruto de cafe coffea arabica l. var. colombia,” *Revista Cenicafe*, vol. 45, pp. 41–50, 1994.
- [2] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second ed., 2004.
- [3] D. G. R. Bradski and A. Kaehler, *Learning OpenCV, 1st Edition*. O’Reilly Media, Inc., first ed., 2008.
- [4] O. Enqvist, C. Olsson, and F. Kahl, “Stable structure from motion using rotational consistency,” tech. rep., Centre for Mathematical Sciences, Lund University, Sweden, 2011.
- [5] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, pp. 91–110, Nov. 2004.
- [6] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.
- [7] Y. Furukawa and J. Ponce, “Accurate, dense, and robust multi-view stereopsis,” in *Computer Vision and Pattern Recognition, 2007. CVPR ’07. IEEE Conference on*, pp. 1–8, June 2007.
- [8] D. Patiño, N. Montes, G. Osorio, and F. Prieto., “Algoritmos de constancia de color para el reconocimiento de objetos.” in *En Memorias del Encuentro de Investigación sobre Tecnologías de Información (EITI-2002)*, p. 145–149, 10 2002.
- [9] P. R. Giraldo, “Desarrollo de una herramienta portátil con visión artificial para la cosecha selectiva del café.” 2014.
- [10] P. Ramos G., J. Sanz U., and C. Oliveros T., “Identificación y clasificación de frutos de café en tiempo teal, a través de la medición de color,” *Cenicafe*, vol. 61, no. 4, pp. 315 – 326, 2010.
- [11] F. R. Ajdadi, Y. A. Gilandeh, K. Mollazade, and R. P. Hasanzadeh, “Application of machine vision for classification of soil aggregate size,” *Soil and Tillage Research*, vol. 162, pp. 8 – 17, 2016.

-
- [12] L. Jiang, A. Koch, S. A. Scherer, and A. Zell, “Multi-class fruit classification using rgb-d data for indoor robots,” in *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 587–592, Dec 2013.
- [13] A. Gollakota and M. B. Srinivas in *2011 Annual IEEE India Conference*, pp. 1–4, Dec 2011.
- [14] M. K. Dutta, A. Singh, and S. Ghosal, “A computer vision based technique for identification of acrylamide in potato chips,” *Computers and Electronics in Agriculture*, vol. 119, pp. 40 – 50, 2015.
- [15] Øystein Sture, E. R. Øye, A. Skavhaug, and J. R. Mathiassen, “A 3d machine vision system for quality grading of atlantic salmon,” *Computers and Electronics in Agriculture*, vol. 123, pp. 142 – 148, 2016.
- [16] B. Zhang, W. Huang, L. Gong, J. Li, C. Zhao, C. Liu, and D. Huang, “Computer vision detection of defective apples using automatic lightness correction and weighted {RVM} classifier,” *Journal of Food Engineering*, vol. 146, pp. 143 – 151, 2015.
- [17] C. Nansen, K. Singh, A. Mian, B. J. Allison, and C. W. Simmons, “Using hyperspectral imaging to characterize consistency of coffee brands and their respective roasting classes,” *Journal of Food Engineering*, vol. 190, pp. 34 – 39, 2016.
- [18] E. M. de Oliveira, D. S. Leme, B. H. G. Barbosa, M. P. Rodarte, and R. G. F. A. Pereira, “A computer vision system for coffee beans classification based on computational intelligence techniques,” *Journal of Food Engineering*, vol. 171, pp. 22 – 27, 2016.
- [19] B. Rodrigues, A. Soares, R. Costa, J. V. Baalen, R. Salvini, F. Silva, M. Caliari, K. Cardoso, T. Ribeiro, A. Delbem, F. Federson, C. Coelho, G. Laureano, and T. Lima, “A feasibility cachaca type recognition using computer vision and pattern recognition,” *Computers and Electronics in Agriculture*, vol. 123, pp. 410 – 414, 2016.
- [20] D. J. Sampson, Y. K. Chang, H. V. Rupasinghe, and Q. U. Zaman, “A dual-view computer-vision system for volume and image texture analysis in multiple apple slices drying,” *Journal of Food Engineering*, vol. 127, pp. 49 – 57, 2014.
- [21] L. Virgen-Navarro, E. J. Herrera-López, R. I. Corona-González, E. Arriola-Guevara, and G. M. Guatemala-Morales, “Neuro-fuzzy model based on digital images for the monitoring of coffee bean color during roasting in a spouted bed,” *Expert Systems with Applications*, vol. 54, pp. 162 – 169, 2016.
- [22] M. Minervini, H. Scharr, and S. A. Tsafaris, “Image analysis: The new bottleneck in plant phenotyping [applications corner],” *IEEE Signal Processing Magazine*, vol. 32, pp. 126–131, July 2015.

-
- [23] G. Moreda, M. Muñoz, M. Ruiz-Altisent, and A. Perdigones, “Shape determination of horticultural produce using two-dimensional computer vision – a review,” *Journal of Food Engineering*, vol. 108, no. 2, pp. 245 – 261, 2012.
- [24] E. Ivorra, A. Sánchez, J. Camarasa, M. Diago, and J. Tardaguila, “Assessment of grape cluster yield components based on 3d descriptors using stereo vision,” *Food Control*, vol. 50, pp. 273 – 282, 2015.
- [25] Y. A. Ohali, “Computer vision based date fruit grading system: Design and implementation,” *Journal of King Saud University - Computer and Information Sciences*, vol. 23, no. 1, pp. 29 – 36, 2011.
- [26] N. El-Bendary, E. E. Hariri, A. E. Hassanien, and A. Badr, “Using machine learning techniques for evaluating tomato ripeness,” *Expert Systems with Applications*, vol. 42, no. 4, pp. 1892 – 1905, 2015.
- [27] Z. Sandoval, F. Prieto, and J. Betancur, “Digital image processing for classification of coffee cherries,” in *Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2010*, pp. 417–421, Sept 2010.
- [28] J. R. Sanz-Urbe, P. J. Ramos-Giraldo, and C. E. Oliveros-Tascon, “Algorithm to identify maturation stages of coffee fruits,” in *World Congress on Engineering and Computer Science 2008, WCECS '08. Advances in Electrical and Electronics Engineering - IAENG Special Edition of the*, pp. 167–174, Oct 2008.
- [29] R. G. Apaza, C. E. Portugal-Zambrano, J. C. Gutiérrez-Cáceres, and C. A. Beltrán-Castañón, “An approach for improve the recognition of defects in coffee beans using retinex algorithms,” in *Computing Conference (CLEI), 2014 XL Latin American*, pp. 1–9, Sept 2014.
- [30] R. H. M. Condori, J. H. C. Humari, C. E. Portugal-Zambrano, J. C. Gutiérrez-Cáceres, and C. A. Beltrán-Castañón, “Automatic classification of physical defects in green coffee beans using cglcm and svm,” in *Computing Conference (CLEI), 2014 XL Latin American*, pp. 1–9, Sept 2014.
- [31] B. Zhang, W. Huang, J. Li, C. Zhao, S. Fan, J. Wu, and C. Liu, “Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review,” *Food Research International*, vol. 62, pp. 326 – 343, 2014.
- [32] A. Djuricic, M. Weinmann, and B. Jutzi, “Potentials of small, lightweight and low cost multi-echo laser scanners for detecting grape berries,” *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-5, pp. 211–216, 2014.

-
- [33] L. Luo, Y. Tang, X. Zou, M. Ye, W. Feng, and G. Li, "Vision-based extraction of spatial information in grape clusters for harvesting robots," *Biosystems Engineering*, vol. 151, pp. 90 – 104, 2016.
- [34] H. N. Patel, R.k.jain, and M.v.joshi, "Automatic segmentation and yield measurement of fruit using shape analysis," *International Journal of Computer Applications*, vol. 45, pp. 19–24, May 2012.
- [35] U. Verma, F. Rossant, I. Bloch, J. Orensanz, and D. Boisgontier, "Shape-based segmentation of tomatoes for agriculture monitoring," in *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, pp. 402–411, March 2014.
- [36] A. Gongal, A. Silwal, S. Amatya, M. Karkee, Q. Zhang, and K. Lewis, "Apple crop-load estimation with over-the-row machine vision system," *Computers and Electronics in Agriculture*, vol. 120, pp. 26 – 35, 2016.
- [37] S. Jay, G. Rabatel, X. Hadoux, D. Moura, and N. Gorretta, "In-field crop row phenotyping from 3d modeling performed using structure from motion," *Computers and Electronics in Agriculture*, vol. 110, pp. 70 – 77, 2015.
- [38] J. Guerrero, M. Guijarro, M. Montalvo, J. Romeo, L. Emmi, A. Ribeiro, and G. Pajares, "Automatic expert system based on images for accuracy crop row detection in maize fields," *Expert Systems with Applications*, vol. 40, no. 2, pp. 656 – 664, 2013.
- [39] J. Romeo, G. Pajares, M. Montalvo, J. Guerrero, M. Guijarro, and J. de la Cruz, "A new expert system for greenness identification in agricultural images," *Expert Systems with Applications*, vol. 40, no. 6, pp. 2275 – 2286, 2013.
- [40] F. Kurtuluş and İsmail Kavdir, "Detecting corn tassels using computer vision and support vector machines," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7390 – 7397, 2014.
- [41] E. R. Hunt, C. S. T. Daughtry, S. B. Mirsky, and W. D. Hively, "Remote sensing with simulated unmanned aircraft imagery for precision agriculture applications," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 4566–4571, Nov 2014.
- [42] H. N. Patel, D. R.K.Jain, and D. M.V.Joshi, "Fruit detection using improved multiple features based algorithm," *International Journal of Computer Applications*, vol. 13, pp. 1–5, January 2011.
- [43] M. Nielsen, D. C. Slaughter, and C. Gliever, "Vision-based 3d peach tree reconstruction for automated blossom thinning," *IEEE Transactions on Industrial Informatics*, vol. 8, pp. 188–196, Feb 2012.

- [44] D. Dey, L. Mummert, and R. Sukthankar, “Classification of plant structures from uncalibrated image sequences,” in *Applications of Computer Vision (WACV), 2012 IEEE Workshop on*, pp. 329–336, Jan 2012.
- [45] R. N. Lati, S. Filin, and H. Eizenberg, “Estimating plant growth parameters using an energy minimization-based stereovision model,” *Computers and Electronics in Agriculture*, vol. 98, pp. 260 – 271, 2013.
- [46] J. Moonrinta, S. Chaivivatrakul, M. N. Dailey, and M. Ekpanyapong, “Fruit detection, tracking, and 3d reconstruction for crop mapping and yield estimation.,” in *ICARCV*, pp. 1181–1186, IEEE, 2010.
- [47] G. Alvarado and H. Ochoa, “Características fenotípicas de compuestos de variedad castillo r en dos ambientes,” *Revista Cenicafé*, vol. 57, no. 2, pp. 100–121, 2006.
- [48] D. C. Brown, “Decentering distortion of lenses,” *Photometric Engineering*, vol. 32, no. 3, pp. 444–462, 1966.
- [49] J. Salvi, X. Armangué, and J. Batlle, “A comparative review of camera calibrating methods with accuracy evaluation,” *Pattern Recognition*, vol. 35, no. 7, pp. 1617 – 1635, 2002.
- [50] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 1330–1334, Nov. 2000.
- [51] A. R. Th, “The levenberg-marquardt algorithm,” 2004.
- [52] S. Baker and I. Matthews, “Lucas-kanade 20 years on: A unifying framework,” *Int. J. Comput. Vision*, vol. 56, pp. 221–255, Feb. 2004.
- [53] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, pp. 273–297, Sept. 1995.
- [54] A. Gidudu, G. Hulley, and T. Marwala, “Image classification using svms: One-against-one vs one-against-all,” *CoRR*, vol. abs/0711.2914, 2007.
- [55] N. Snavely, S. M. Seitz, and R. Szeliski, “Photo tourism: Exploring photo collections in 3d,” *ACM Trans. Graph.*, vol. 25, pp. 835–846, July 2006.
- [56] N. Snavely, S. M. Seitz, and R. Szeliski, “Modeling the world from internet photo collections,” *Int. J. Comput. Vision*, vol. 80, pp. 189–210, Nov. 2008.
- [57] P. Ramos, F. Prieto, and C. Oliveros, “Semiautomatic annotation system to generate ground truth in image sequence of agricultural sceneries,” in *6th International Workshop Applications of Computer Image Analysis and Spectroscopy in Agriculture*, 2016. Aarhus - Denmark.

-
- [58] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [59] E. Osuna, R. Freund, and F. Girosi, “Support vector machines: Training and applications,” tech. rep., Cambridge, MA, USA, 1997.
- [60] L. Luo and X. Chen, “Integrating piecewise linear representation and weighted support vector machine for stock trading signal prediction,” *Applied Soft Computing*, vol. 13, no. 2, pp. 806 – 816, 2013.
- [61] P. Ramos, F. Prieto, E. Montoya, and C. Oliveros, “Automatic fruit count on coffee branches using computer vision,” *Computers and Electronics in Agriculture*, vol. 137, pp. 9 – 22, 2017.