



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

ALGORITMO PARA LA GENERACIÓN DE ORTOMOSAICOS A PARTIR DE  
IMÁGENES AÉREAS TOMADAS POR DRONES EN LA AGRICULTURA

MEMORIA PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL ELÉCTRICO

ALVARO RODRIGO GÓMEZ ORTEGA

PROFESOR GUÍA:  
CLAUDIO ESTÉVEZ MONTERO

MIEMBROS DE LA COMISIÓN:  
CHRISTOPHER POPE SCHWARTZ  
FRANCISCO GALDAMES GRÜNBERG

Este trabajo ha sido parcialmente financiado por Telefónica I+D Chile

SANTIAGO DE CHILE

2018

RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO  
POR: ALVARO RODRIGO GÓMEZ ORTEGA  
FECHA: 2018  
PROF. GUÍA: CLAUDIO ESTÉVEZ MONTERO

## ALGORITMO PARA LA GENERACIÓN DE ORTOMOSAICOS A PARTIR DE IMÁGENES AÉREAS TOMADAS POR DRONES EN LA AGRICULTURA

La mejora en los vehículos aéreos no tripulados, las cámaras digitales y la fotogrametría, ha permitido que en la agricultura se tengan varios beneficios como: detectar el estado hídrico de las plantas, revelar el estrés nutricional en cultivos, localizar de manera temprana enfermedades o plagas y calcular índices relativos a la calidad de las plantaciones. Dichos beneficios han hecho posible mejorar la gestión del agricultor y su toma de decisiones sobre sus terrenos. Gran parte de esto puede ser logrado superponiendo imágenes aéreas verticales de un terreno agrícola donde se mantengan las dimensiones geométricas. Bajo esta mirada, los objetivos generales de esta memoria consisten en comprender, analizar, diseñar, implementar y evaluar el desarrollo de un algoritmo que genere ortomosaicos de manera automática para fotografías tomadas por drones de manera vertical que incluyan un sistema de posicionamiento global (GPS) y una unidad de medición inercial (IMU).

Para lograr el objetivo se propone una metodología modular que comienza con las imágenes en bruto, tal que utilizando su información geoespacial y la inclinación de la cámara agilice la identificación de la información considerada perpendicular para eliminar los sectores que no lo sean y estimar cuáles fotografías debieran estar solapadas. Luego, se busca con mayor certeza entre estos conjuntos cuales comparten información identificando una serie de puntos claves, los que serán utilizados para realizar distintas homografías a cada imagen y así superponerlas de manera adecuada conservando las proporciones geométricas creando así una primera versión del ortomosaico. Después se realizan mejoras en su aspecto, haciendo un ajuste de intensidades entre las imágenes y finalmente mezclando las fotografías para que los bordes de cada una dejen de ser perceptibles.

Con la metodología diseñada se consiguen ortomosaicos de buen aspecto, con una baja distorsión espacial, y un orden de cómputo lineal en casi todas sus fases. Además, es un algoritmo que resulta ser escalable y concretamente aplicable al contexto real de la industria de la agricultura actual, mostrando así que los resultados obtenidos indican que el procedimiento contribuye a los objetivos planteados. En resumen, el presente trabajo muestra un caso sobre cómo las tecnologías actuales pueden ser utilizadas en la industria de la agricultura actual, permitiendo generar una visión completa de un sector de interés en una sola imagen, mediante fotografías no necesariamente de alta resolución, siempre y cuando, estas sean tomadas de manera perpendicular a la superficie del terreno.



*A mis Ávaros del pasado, a mis Ávaros del presente y a mis Ávaros del futuro.*



# Agradecimientos

Es fácil agradecer cuando los ojos se encuentran inyectados de euforia, producido por una alteración del pensamiento donde la sangre se encuentra a una temperatura tal, que no cabe duda la excitación que recorre por todo el cuerpo. Cuando se puede trabajar y producir sin necesidad de reposo alguno, envuelto hasta el último rincón por una excesiva confianza, capaz de generar una total desinhibición en el actuar. En el momento en que se tiene un desorden bioquímico adictivo, donde lo único que se percibe al mirar un espejo es una inigualable grandeza. En la ocasión que al reflexionar sobre los objetivos que se tienen y mutar personalmente en un depredador de los mismos, en el que cada paso dado es realizado con la sed del cazador.

Es grato poder decidir si agradecer estando tranquilo, con el corazón latiendo a un perfecto compás y los pulmones realizando un vaivén magistral. En el instante en que la gran mayoría de dimensiones involucradas en la existencia parecen estar en orden y aquellas que no, coexistiendo de manera armónica dentro del caos. Cuando todo parece avanzar a un ritmo regular, a la velocidad perfecta, donde cada oscilación es amortiguada por la tranquilidad de estar viviendo.

Es fácil desagradecer cuando los ojos han perdido todo brillo y no aguantan el momento de cerrarse por días, adjuntos a un cuerpo vacío mantenido en funcionamiento gracias a pequeños seres interesados por sí mismos. Cuando el sentimiento de inutilidad y la desesperanza se apoderan del propio pensar, siendo imposible el goce de las actividades que alguna vez se realizaban con tantas ganas. En el instante en que la propia habitación parece el lugar menos incómodo del universo. Cuando no se elige y aparece semejante miseria, aquella que por mucho que se intente eliminar, lo único que hace es seguir carcomiendo lo poco que queda.

¿Es este trabajo merecedor de una página de agradecimientos? Al parecer siempre dependerá de cuál Álvaro esté recordando este momento.



# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
<b>2. Marco Teórico</b>	<b>4</b>
2.1. Teledetección	4
2.1.1. Energía electromagnética	5
2.1.2. Sensores electromagnéticos	7
2.2. Fotografía aérea	8
2.2.1. Vehículos aéreos no tripulados	9
2.3. Fotogrametría	10
2.3.1. Geometría ideal de una fotografía aérea	11
2.3.2. Geometría real de una fotografía aérea	12
2.4. Procesamiento imágenes	13
2.4.1. Transformación de imágenes	15
2.4.2. Índices de vegetación para la agricultura de precisión	15
2.5. Ortomosaicos	17
2.5.1. SIFT para encontrar puntos claves	18
2.5.2. Matriz de Homografía para superponer imágenes	19
2.5.3. RANSAC para la eliminación de emparejamientos erróneos	22
2.5.4. Compensación de ganancia para ajustar intensidades de imágenes solapadas	22
2.5.5. Mezcla de imágenes en múltiples bandas para reducir efectos no deseados	24
<b>3. Metodología</b>	<b>27</b>
3.1. Geometría de la cámara en la agricultura	27
3.2. Preparación de las fotografías	29
3.3. Búsqueda de puntos comunes entre fotografías	31
3.3.1. Detección de puntos claves	31
3.3.2. Emparejamiento de puntos claves	33
3.3.3. Eliminación de emparejamientos erróneos	33
3.4. Mejoramiento de la matriz de homografía	35
3.5. Mejoras al aspecto del ortomosaico	37
3.5.1. Ajuste de intensidades de las fotografías con ganancias variables	37
3.5.2. Mezcla de imágenes en múltiples bandas	38
3.6. Generación de métricas para medir el desempeño del algoritmo	39
3.6.1. Artefactos visuales	39
3.6.2. Precisión geométrica	39



3.6.3. Costo computacional . . . . .	40
<b>4. Resultados</b>	<b>41</b>
4.1. Criterio de perpendicularidad . . . . .	41
4.2. Información del GPS e IMU en la búsqueda de puntos comunes . . . . .	43
4.3. Estimación de las matrices de homografía con respecto al plano de referencia	46
4.4. Cambios de aspecto del ortomosaico . . . . .	49
<b>Conclusión</b>	<b>55</b>
<b>Bibliografía</b>	<b>57</b>

# Índice de Tablas

4.1. Errores espaciales obtenidos en ambos ortomosaico finales. . . . .	55
---	----

# Índice de Ilustraciones

2.1. Regiones del espectro electromagnético [1]. . . . .	5
2.2. Soporte para fotografía aérea [2]. . . . .	7
2.3. Funcionamiento cámaras digitales [2]. . . . .	8
2.4. Tipos de fotografías aéreas [3]. . . . .	8
2.5. Algunos tipos de vehículos aéreos no tripulados [4]. . . . .	9
2.6. Bloque de fotografías aéreas con 60 % de solapamiento, en tres líneas de vuelo con 25 % de solapamiento. Este diseño es ideal para asegurar una cobertura completa y con la redundancia suficiente para aplicaciones posteriores [3]. . .	10
2.7. Fotografía vertical tomada sobre terreno plano. El eje óptico, que interseca el plano de la imagen en su centro $o$ , encuentra el suelo en un ángulo recto en el punto principal $P$ . Obsérvese que las distancias respectivas entre todos los puntos son las mismas en el suelo y en la imagen [3] . . . . .	11
2.8. Fotografía vertical tomada sobre terreno variable. La elevación del punto principal $P$ determina el plano horizontal del dato local. Los puntos situados en este plano permanecen sin distorsión, mientras que los puntos arriba o abajo se desplazan radialmente con respecto al centro de la imagen. Observe que las distancias horizontales $D_1 - D_4$ son las mismas en el espacio de objetos, pero no en la imagen. [3] . . . . .	12
2.9. Fotografía oblicua tomada sobre un terreno completamente plano. La línea vertical que atraviesa el centro $L$ en perspectiva interseca el plano de la imagen en el punto nadir fotográfico $n$ y se encuentra con el suelo en ángulo recto en el punto nadir $N$ . El ángulo entre la línea nadir y el eje óptico es el ángulo nadir $v$ . Obsérvese que las distancias $D_i$ entre todos los puntos en el suelo son iguales, pero las distancias $d$ correspondientes en la imagen varían continuamente [3] . . . . .	13
2.10. Ejemplos de defectos que pueden tener las imágenes. Ejemplo A: Aberraciones radiales. Ejemplo B: Vignetting. Ejemplo C: Aberraciones cromáticas. Ejemplo D: Ruido. [5] [6] [7] [8] . . . . .	14
2.11. Fotografía de un terreno agrícola representado por el índice de vegetación NDVI [9]. . . . .	16
2.12. Mosaico rugoso formado a partir de imágenes aéreas usando solamente los datos del GPS y de la IMU [10]. . . . .	17
2.13. Una homografía puede tener distintos tipos de transformaciones visuales según los grados de libertad que esta tenga. El primer caso es la original, luego la traslación con 2 grados de libertad, la transformación euclidiana para 3 grados de libertad, transformación afín para 6 grados de libertad y homografía para 8 [11]. . . . .	19

2.14.	A la izquierda se tiene un conjunto de datos con distintos valores atípicos. En la derecha el algoritmo RANSAC encuentra un modelo que explica la mayoría de los datos y permite identificar cuáles son los valores atípicos [12]. . . . .	22
2.15.	Ejemplo de mezcla de imágenes. Figura A y B, corresponden a las imágenes a mezclar. Figura C es el caso en que se pone una mitad sobre la otra sin técnica alguna de mezcla. Figura D es el caso en que se ocupa una técnica de mezcla, haciendo que los bordes de ambas dejen de percibirse [13]. . . . .	25
3.1.	Representación de la geometría de una cámara al tomar una fotografía sobre terreno agrícola, donde el objeto verde representa un árbol captado completamente por la misma. . . . .	28
3.2.	Representación de tres casos que se pueden dar entre dos fotografías al momento de verificar el solapamiento entre ellas. Caso A: Una de las fotografías posee dos de sus vértices dentro de otra. Caso B: Las fotografías no se encuentran solapadas. Caso C: Una de las fotografías posee un vértice dentro de otra. Cualquier otro caso de solapamiento no es considerado, debido a la naturaleza de la toma de fotografías con drones. . . . .	30
3.3.	Muestra de los dos casos que se pueden dar cuando se quiere determinar si algún vértice de un rectángulo se encuentra dentro de otro. Caso A: Si un vértice de un rectángulo se encuentra dentro de otro, al trazar cualquier rayo desde el punto intersecará en una sola arista del rectángulo. Caso B: Si un vértice se encuentra fuera del rectángulo, al trazar cualquier rayo desde el punto intersecará con ninguno o con dos aristas del rectángulo. . . . .	31
3.4.	Par de imágenes en que se quiere verificar si están solapadas. Con el algoritmo SIFT se encuentra un conjunto de puntos repartido a lo largo de cada fotografía [14]. . . . .	32
3.5.	Emparejamiento de puntos claves entre dos imágenes basado en descriptores locales [15]. . . . .	33
3.6.	Emparejamientos mantenidos luego de realizar el algoritmo RANSAC sobre los emparejamientos encontrados. Con esto se borran la mayor parte de los falsos positivos emparejados [16]. . . . .	34
3.7.	Dos imágenes que realizando una transformación homográfica adecuada quedan superpuestas [16]. . . . .	34
3.8.	Representación gráfica de la diferencia entre la proyección del emparejamiento $i$ -ésimo sobre el plano de referencia. Caso A: Se muestran ambas fotografías involucradas en el emparejamiento $i$ -ésimo con sus puntos correspondientes. Caso B: Se muestra el caso en que la matriz de homografía hacia el plano de referencia fuese perfecta, calzando en el mismo pixel ambos puntos, siendo el error de proyección $\ e_i\ ^2 = 0$ . Caso C: Se muestra el caso en que las proyecciones no logran encontrarse en el mismo pixel, obteniéndose un error $\ e_i\ ^2 > 0$ . . . . .	36
3.9.	Diferencia de los resultados de un mismo mosaico cuando se realiza un ajuste de intensidades de las fotografías con ganancias variables [14]. . . . .	37
3.10.	Comparación de la distribución obtenida entre una función lineal (imagen izquierda) y una no lineal (imagen derecha). La última corresponde a una función que genera una distribución elipsoidal proporcional a las correspondencias de la imagen. . . . .	38

4.1. Ajuste que recibe una fotografía urbana (imagen de la izquierda) al momento de redimensionarla para que cada uno de sus pixeles cumpla con el criterio de perpendicularidad (imagen de la derecha). . . . .	42
4.2. Ajuste que recibe una fotografía agrícola (imagen de la izquierda) al momento de redimensionarla para que cada uno de sus pixeles cumpla con el criterio de perpendicularidad (imagen de la derecha). . . . .	42
4.3. Emparejamiento realizado entre dos imágenes solapadas (fotografía de arriba), donde los emparejamientos mal realizados quedan completamente eliminados con el algoritmo RANSAC como se ve en la fotografía de abajo. . . . .	43
4.4. Emparejamiento realizado entre dos imágenes que no están solapadas (Fotografía de arriba). Los calces realizados son todos falsos positivos, los que con el algoritmo RANSAC quedan completamente eliminados, concluyendo que estas fotografías no comparten información de una escena (fotografía de abajo). . .	44
4.5. Emparejamiento realizado para dos fotografías muy similares que se encuentran solapadas. En la fotografía de arriba se puede ver la gran cantidad de emparejamientos, sin una dirección clara a simple vista, sin embargo, tras el algoritmo RANSAC, se logran encontrar un par de calces que explican el solapamiento entre ambas imágenes (fotografía de abajo). . . . .	44
4.6. Gráfico que muestra el tiempo promedio de procesamiento que toma emparejar cuatro imágenes según distintas escalas aplicadas a las imágenes. . . . .	45
4.7. Gráfico que muestra el tiempo promedio de procesamiento que toma emparejar distintas cantidades de imágenes para una escala fija de 0.1. . . . .	46
4.8. Ortomosaico defectuoso debido a la baja calidad y/o cantidad de emparejamientos encontrados en las etapas previas, junto a una falta de mayor procesamiento para llegar al óptimo de la estimación de homografías. . . . .	47
4.9. Imagen original que es deformada en el ortomosaico de la Figura 4.8. . . . .	47
4.10. Ortomosaico del Escenario A obtenido de buen aspecto según usuario experto.	48
4.11. Ortomosaico del Escenario B obtenido de buen aspecto según usuario experto.	49
4.12. Escenario A (izquierda) al que se le aplica un ajuste de intensidades en base a ganancias variables (derecha). . . . .	50
4.13. Escenario B (izquierda) al que se le aplica un ajuste de intensidades en base a ganancias variables (derecha). . . . .	50
4.14. Escenario A al realizar una mezcla de imágenes en múltiples bandas. . . . .	51
4.15. Escenario B al realizar una mezcla de imágenes en múltiples bandas. . . . .	52
4.16. Escenario A al eliminar sus bordes tras realizar los ajustes de apariencia. . .	53
4.17. Escenario B al eliminar sus bordes tras realizar los ajustes de apariencia . . .	54

# Capítulo 1

## Introducción

### Motivación

Los vertiginosos avances en las tecnologías han permitido que los agricultores puedan obtener altas cantidades de datos específicos de sus campos, con el objetivo de mejorar la toma de decisiones en las prácticas agrícolas. Esto generó que una nueva área comenzara a estudiarse, la agricultura de precisión, definida como la gestión de la variabilidad espacial y temporal en la agricultura para mejorar el rendimiento económico obtenido de los cultivos, reducir el impacto ambiental, y mejorar las condiciones sociales. Dicha ciencia tiene como actividades principales la adquisición, procesamiento y obtención de información a partir de los datos recolectados. Para la realización de dichas actividades se tiene un gran número de sensores, capaces de capturar y almacenar variados tipos de medidas gracias a las tecnologías de la información, y una amplia gama de algoritmos y estudios relacionados [17].

Es en este contexto que los vehículos aéreos no tripulados han tomado un peso importante en la agricultura ya que, sin el uso de esta tecnología, el agricultor deba desplazarse por el campo determinando visualmente las deficiencias que puedan existir en el cultivo, y generalmente con baja precisión. La ventaja del dron es que puede volar sobre las tierras y obtener la información del estado de las plantas con mayor fidelidad, debido a la posibilidad de capturar fotografías de la plantación en variados espectros. Todo esto permite analizar diferentes problemas que afectan al desarrollo de los cultivos a partir de fotografías, como por ejemplo [18]:

- Detección del estado hídrico de las plantas: Mediante sensores térmicos es posible monitorizar el estrés hídrico en los cultivos, ya que éste fenómeno provoca el cierre de los estomas, reduciendo la transpiración y aumentando la temperatura de las hojas.
- Detección de estrés nutricional en cultivos: Determinando el contenido de clorofila de las plantas se puede obtener la concentración de nitrógeno en las hojas, algo posible de medir con imágenes de variados espectros, lo que permite optimizar el uso de fertilizantes.
- Detección temprana de enfermedades y plagas: La asistencia de imágenes multiespec-

trales permite detectar el estrés en los cultivos causados por la presencia de plagas y enfermedades, generando variados mapas con el propósito de evaluar de mejor forma la situación.

- Índices relativos a calidad en cultivos: Mediante la creación de mapas representativos se puede generar información muy valiosa, impactando en la rentabilidad del campo.

Para que el análisis de las plantaciones sea utilizable de manera eficiente y adecuada es necesario combinar todas las fotos, con el objetivo de tener un mapa completo de la plantación, y que el agrónomo a cargo de la tierra pueda tener una vista general del estado de ésta. Con este procedimiento es posible observar qué sectores del terreno se encuentran más saludables que otros, para realizar acciones correctivas sobre las áreas que lo necesiten. Para realizar este procedimiento no basta con superponer las fotografías unas sobre otras, ya que éstas podrían ser tomadas a diferentes alturas con el dron, a distintos ángulos, con variaciones de iluminación, entre otras muchas situaciones que complican el proceso. Por todo lo descrito anteriormente, es que en este trabajo se diseña y desarrolla un algoritmo para la generación de mapas sin distorsiones geométricas y perpendiculares a la superficie (ortomosaicos) para la agricultura.

## Hipótesis

El uso de fotografías aéreas verticales tomadas por drones que posean un sistema de posicionamiento global (GPS) y una unidad de medición inercial (IMU) sobre terrenos agrícolas permite la creación de ortomosaicos tal que, si son generados automáticamente y poseen buen aspecto, pocas distorsiones espaciales y bajos tiempos de cómputo, pueden utilizarse para variados propósitos en lo que respecta a la toma de decisiones sobre los cultivos en estudio, siendo un aporte a la agricultura de precisión.

## Objetivos generales

Los objetivos generales del presente trabajo consisten en comprender, analizar, diseñar, implementar y evaluar, el desarrollo de un algoritmo que genere ortomosaicos de forma automática para fotografías tomadas por drones de manera vertical que incluyan un sistema de posicionamiento global (GPS) y una unidad de medición inercial (IMU). Además, esta metodología debe funcionar sin la intervención de un usuario para la configuración de parámetros iniciales, exceptuando la ubicación del directorio en que se encuentran las imágenes. Finalmente, el resultado obtenido debe poseer las características necesarias para poder tomar mejores decisiones en la agricultura y aplicaciones afines.

## Objetivos específicos

Para conseguir los objetivos generales se definen los siguientes objetivos específicos:

1. Identificar los conceptos relacionados a la teledetección, la fotografía aérea, y la fotogrametría.
2. Relacionar los conceptos de la fotogrametría con la fotografía aérea tomada por vehículos aéreos no tripulados.
3. Enlistar los problemas de mayor impacto que tienen las imágenes aéreas independiente de la cámara con que se tomen.
4. Explicar algunas de las aplicaciones y ventajas que tiene la superposición de imágenes en distintas bandas para la agricultura de precisión.
5. Definir lo que es un ortomosaico y las características que éstos deben tener para generar valor en la toma de decisiones en la agricultura.
6. Identificar, resumir, y justificar distintas herramientas encontradas en la literatura que pueden ser usadas para generar un ortomosaico.
7. Inventar, modificar, esquematizar y detallar un flujo lineal para la generación de un ortomosaico de manera automática, para fotografías tomadas verticalmente, basándose en la información recogida.
8. Generar variadas métricas para medir el desempeño del algoritmo creado, en base a las aplicaciones que éste tendría en la agricultura de precisión.
9. Programar el algoritmo diseñado para la generación de un ortomosaico de manera automática en el lenguaje Matlab para un conjunto de imágenes aéreas.
10. Aprender y criticar los resultados obtenidos en base a las métricas de desempeño generadas.
11. Concluir y valorar si el algoritmo diseñado es utilizable en el campo de la agricultura de precisión.
12. Proponer mejoras para trabajos posteriores que puedan reflejarse en el desempeño del algoritmo desarrollado.

## Alcances

En esta memoria de título se trabaja con imágenes RGB que posean sus datos GPS e IMU correspondientes, ya sea como metadatos o en un archivo de texto aparte. Además, las fotografías deben estar tomadas de manera vertical sobre un terreno agrícola. Finalmente, para la programación del algoritmo, se considera el uso de Matlab y los toolboxes llamados Computer Vision System Toolbox, y Optimization Toolbox.



# Capítulo 2

## Marco Teórico

Para entender este trabajo de título sin la necesidad de recurrir a fuentes externas, en este capítulo se mencionan y explican todos aquellos tópicos relacionados con el tema principal. Para ello se comienza describiendo cómo la información referente a la vegetación cultivada en distintas hectáreas agrícolas puede ser almacenada en una imagen, y luego se explica el procedimiento que permite trabajar estas fotografías para convertirlas en un ortomosaico.

### 2.1. Teledetección

La ciencia dedicada a obtener información de un objeto o área de la superficie de la tierra sin estar en contacto directo con éste es llamada teledetección. La gran mayoría de seres vivos realizan esta tarea utilizando sus sentidos, como la vista, el gusto, el olfato o el tacto, por lo que la detección remota es algo que existe desde antes de que fuese concebido el concepto. Es una ciencia multidisciplinaria que integra una serie de tecnologías y disciplinas como óptica; espectroscopía; computación; electrónica; telecomunicaciones; entre otras, para actuar como un solo sistema completo [19].

Existen varias etapas en el proceso de teledetección, y cada una de ellas es importante para una operación exitosa: Adquisición de datos; transmisión de éstos; procesamiento; y análisis. La gran mayoría de dispositivos utilizados en esta área y que captan información sobre un objeto (sensores), detectan y graban la transmisión de energía electromagnética reflejada o irradiada de su superficie, donde se aprovecha que cada uno devuelve distintas cantidades y bandas del espectro electromagnético dependiendo de las propiedades del material; rugosidad de la superficie; ángulo de incidencia; intensidad; y longitud de onda de la energía irradiada. A continuación, se explicará en qué consiste la energía electromagnética, las propiedades que posee, y de dónde surge la respuesta particular de cada objeto [19].

### 2.1.1. Energía electromagnética

La energía, una medida de la habilidad para realizar trabajo, puede estar presente en muchas formas y tiene la capacidad de ser transformada de un tipo a otro. Una de ellas es la energía electromagnética, una forma importante de transportar energía, generada a partir de partículas cargadas en movimiento, donde una variación en un campo magnético introducirá un cambio en un campo eléctrico o viceversa, siempre relacionados. Este tipo de energía viaja en ondas a la velocidad de la luz, sin requerir de un medio para propagarse, y abarca un amplio rango desde las ondas de radio hasta los rayos gamas. El Sol es una fuente de energía que abarca todo este rango y su radiación electromagnética atraviesa la atmósfera del planeta constantemente, interactuando con las partículas de la Tierra. El ojo humano puede detectar solo una porción de esta extensión llamado luz visible, y para detectar las otras porciones es necesario utilizar otro tipo de sensores lo que será comentado más adelante en este capítulo [20].

La energía electromagnética se distribuye en un intervalo casi infinito de frecuencias, es decir, la cantidad de veces en que sucede un movimiento periódico, llamado espectro electromagnético. Este se ha dividido en variadas regiones, sin embargo, no tienen fronteras rígidas entre las extensiones adyacentes y dependen en buena medida en la capacidad tecnológica de poder producirlas o detectarlas. En la Figura 2.1 se muestran los rangos aproximados de frecuencias de las diferentes regiones del espectro electromagnético [21] [22].

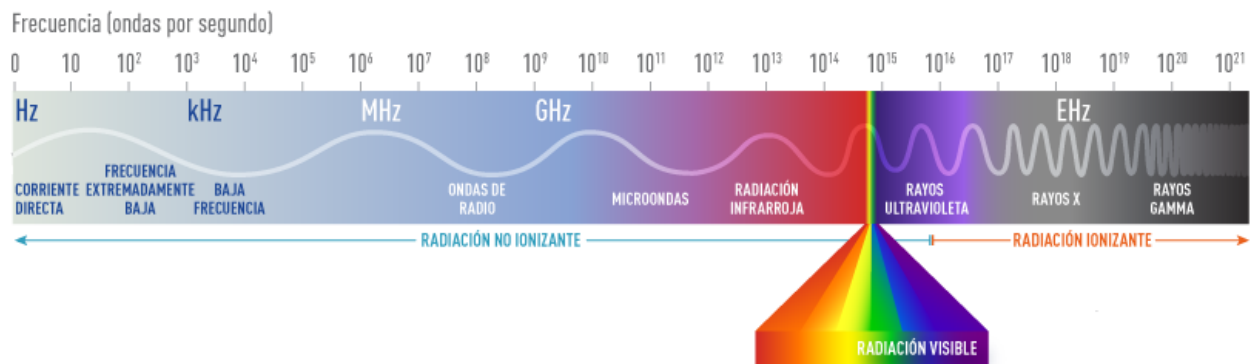


Figura 2.1: Regiones del espectro electromagnético [1].

En lo que respecta a la teledetección en la agricultura, no son todas las bandas del espectro electromagnético las que poseen información de interés. Aquellas más utilizadas por los investigadores del área corresponden a las ondas infrarrojas; la luz visible; y las ondas ultravioletas [20]:

- **Ondas infrarrojas:** Esta región del espectro es dividido en infrarrojo cercano; infrarrojo medio; e infrarrojo lejano. La región entre los 8 y 15 micrones es conocida como infrarrojo termal, ya que esta longitud de onda es usada para estudiar la energía térmica irradiada de los cuerpos. Por otra parte, se puede captar la energía infrarroja como calor. Algunos objetos están tan calientes que emiten también luz visible, como el fuego, y otros cuerpos a temperaturas más bajas solo emiten ondas infrarrojas. Los científicos pueden estudiar la salud de la vegetación y la composición del suelo por la radiación infrarroja cercana, ya que la vegetación saludable absorbe la luz azul y roja

(luz visible) para estimular la fotosíntesis y crear la clorofila. Una planta con mayor clorofila reflejará una mayor cantidad de energía infrarroja que una planta no saludable, pudiendo obtener información sobre su productividad.

- **Luz visible:** Toda la radiación electromagnética es luz, pero el humano puede solo ver una pequeña porción de ésta, la que llamamos luz visible. Las células del ojo llamadas conos actúan como receptores sintonizados para esta estrecha banda. Cuando el espectro completo de la luz visible viaja a través de un prisma, los largos de onda se separan en los colores del arcoíris porque cada color tiene un largo de onda distinto, el violeta tiene la longitud de onda más corta, cerca de los 380 nanómetros, y la roja cerca de los 700 nanómetros.
- **Ondas ultravioletas:** Tienen una longitud de onda más corta que la luz visible. El sol es una fuente del espectro completo de radiación ultravioleta, que es comúnmente dividida en UV-A, UV-B y UV-C. Los rayos UV-C son los más dañinos y son casi completamente absorbidos por la atmósfera terrestre. Por otra parte, los rayos UV-B son los que producen el bronceado y principal fuente de cambios celulares.

Para entender cómo los sensores captan la información de una superficie gracias a las ondas electromagnéticas, es necesario comprender cómo la luz del Sol interactúa con la materia, explicando así, lo que permite que cada objeto tenga una respuesta diferente. Cuando una onda electromagnética penetra algún cuerpo, su campo electromagnético interactúa con el campo localizado de los átomos que lo forman, cuyo resultado es que cuando la luz emerge de la materia, sus características y propiedades podrían no ser las mismas. Los efectos que causa la materia que interactúa con este tipo de energía, dependen de la fuerza de su campo electromagnético; su longitud de onda; la materia en cuestión; influencias externas como la temperatura, presión, y otros. En la materia densa, la luz es absorbida rápidamente dentro de las primeras pocas capas de átomos, ya que no emerge de ésta, y en contraste, algunos tipos de materia no absorben luz completamente, llamadas materias ópticamente transparentes, que permiten a la luz propagarse a través de estos y emerger del medio. A continuación, se examinará la interacción de la luz con este tipo de materiales [23] [24]:

- **Absorción:** Es el fenómeno con el que la intensidad de una onda disminuye debido a efectos disipativos del medio por el que se propaga, produciendo una reducción de la energía transportada.
- **Reflexión** Corresponde al cambio de dirección de la propagación de una onda que ocurre en el mismo medio en que se propagaba, luego de incidir sobre la superficie de un medio distinto, donde el ángulo de incidencia es igual al reflejado.
- **Refracción:** Es el cambio de dirección de la propagación de una onda que ocurre tras pasar a otro medio, en el que se propaga con distinta velocidad. La onda incidente y la refractada están en el mismo plano, y el ángulo de cada una se relacionan con la velocidad por la ley de Snell.
- **Difracción:** Es un fenómeno característico de las ondas que permite bordear obstáculos o difundirse a través de una rendija por la distorsión que le ocurre, reproduciéndose y propagándose en todas las direcciones por dicho orificio. Este fenómeno ocurre cuando el tamaño del orificio o el obstáculo son comparables al de la longitud de onda.
- **Dispersión:** Es el fenómeno por el que las ondas son forzadas a desviarse de su trayectoria recta por uno o más caminos por la localización de aberraciones no uniformes

en el medio por el que está viajando.

### 2.1.2. Sensores electromagnéticos

Para captar la información en la agricultura de forma remota, existen una gran cantidad de sensores, tanto pasivos como activos, capaces de detectar energía electromagnética de distintas regiones del espectro. Los sensores pasivos captan la energía reflejada o emitida de un cuerpo objetivo; y los sensores activos transmiten una señal y reciben la respuesta reflejada. Aquellos más utilizados actualmente para esta aplicación son las cámaras aéreas, un sensor pasivo montado en la parte baja del vehículo aéreo teniendo libre la abertura del lente hacia el plano de la tierra que se desee fotografiar, capaces de crear imágenes en distintos rangos de luz (Figura 2.2). Existen una gran cantidad de cámaras aéreas en el mercado y la gran mayoría tienen elevados precios por poseer una construcción rígida, controlada y meticulosa. Estos sistemas incluyen tanto cámaras analógicas como digitales, donde las digitales no necesitan pasar por un proceso intermedio como procesar las cintas o imprimirlas, a diferencia de las analógicas. Por otra parte, la cámara digital en lugar de captar y almacenar las fotografías en una película química como las analógicas, utiliza elementos de la fotografía digital para generarlas y guardarlas por separado [25].



Figura 2.2: Soporte para fotografía aérea [2].

Las cámaras digitales para capturar las imágenes convierten la información de ésta en una señal eléctrica, donde un sensor es utilizado como dispositivo de conversión, sin embargo, el sensor usado en este tipo de sistema no sirve para almacenar la información. Este sensor es un filtro de mosaico para detectar información de color, dispuesto en la superficie de recepción de la escena y recibe la luz incidente en cada pixel a través de un conjunto de microlentes que ayudan a condensarla en cada uno de ellos. Estos sensores pueden variar en capacidad y resolución, lo que afectará la claridad y el nivel de detalle de la imagen. Todo esto permitirá guardar las imágenes en una memoria, analizarlas y procesarlas de forma computacional (Figura 2.3), lo que hace que hoy sean las cámaras más utilizadas en este tipo de aplicaciones [26].

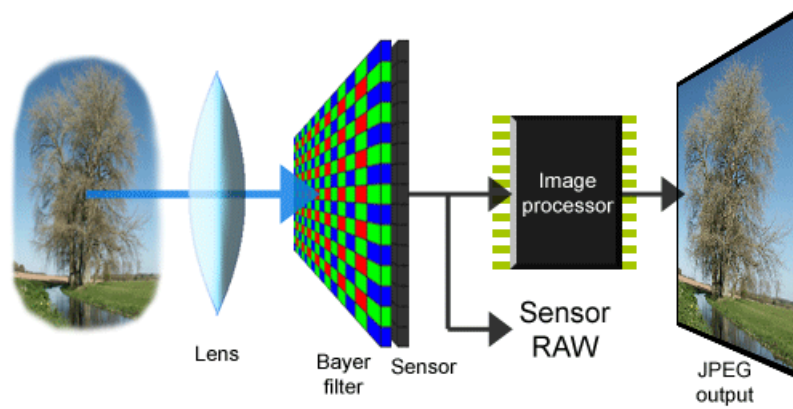


Figura 2.3: Funcionamiento cámaras digitales [2].

## 2.2. Fotografía aérea

Un formato para hacer teledetección es tomando fotografías de alguna parte de la tierra desde el aire, usando cámaras livianas originalmente pensadas para usarlas tanto a mano como con un trípode por amateurs y profesionales, por lo que carecen de la fidelidad tanto geométrica como espacial que poseen las cámaras más grandes y costosa. Además, están diseñadas para tomar fotografías desde una altura menor a 500 metros, y tienen el objetivo de ser altamente portables, haciendo que este formato, conocido como fotografía aérea de formato pequeño (SFAP), pueda ser utilizado para una serie de aplicaciones. La implementación de técnicas de análisis de estas imágenes puede hacerlas una misma persona, permitiendo mayor flexibilidad y especialización, motivo por el que es tan llamativo para científicos y organizaciones. El formato aplicado descrito es muy útil para investigaciones detalladas de condiciones del ambiente en la superficie terrestre siendo aplicada en varias áreas, desde la geociencia hasta la arqueología, investigación de escenas de crímenes y el desarrollo inmobiliario. Las fotografías aéreas pueden ser tomadas en tres vistas determinadas relativas a la superficie de la tierra, determinada por la inclinación del lente con respecto al horizonte (ángulo de depresión): posición oblicua alta (menor a  $20^\circ$ ) (Figura 2.4 A); posición oblicua baja (entre  $20^\circ$  y  $87^\circ$ ) (Figura 2.4 B); y posición vertical (entre  $87^\circ$  y  $90^\circ$ ) (Figura 2.4 C), cada una utilizada según el objetivo que se desee [3].



Figura 2.4: Tipos de fotografías aéreas [3].

### 2.2.1. Vehículos aéreos no tripulados

Los vehículos aéreos no tripulados (VANT), también conocidos como drones, son vehículos aéreos motorizados y no tripulados de variados tipos, controlados remotamente de forma manual, semi automática, o automática (Figura 2.5). Al poder equipar estas plataformas con una herramienta de medición de tamaño pequeño o mediano como las descritas anteriormente, es posible realizar fotografía aérea de formato pequeño sobre esta plataforma. Además, la tecnología actual permite registrar y rastrear la posición y orientación de los sensores en un sistema de coordenadas tanto global como local. Es así como los drones han permitido una gran cantidad de nuevas aplicaciones de corta distancia a la teledetección, casi de tiempo real, y con alternativas de bajo costo a las tecnologías relacionadas con vehículos aéreos tripulados [27].

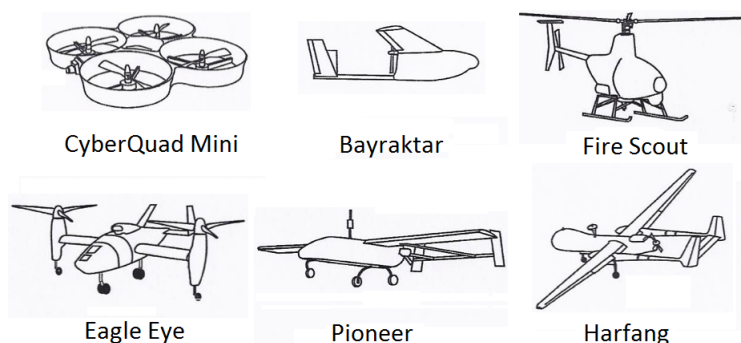


Figura 2.5: Algunos tipos de vehículos aéreos no tripulados [4].

Las ventajas de estas plataformas por sobre las tripuladas, son que los VANT pueden ser usados en situaciones de alto riesgo sin poner en peligro la vida de un humano en áreas inaccesibles o de difícil tránsito, a baja altura y cercano al objeto de interés, algo que anterior a esta tecnología se creía imposible. Incluso, bajo condiciones climáticas nubosas o lluviosas, la adquisición de datos con esta plataforma sigue siendo posible cuando la distancia al objeto permite volar bajo las nubes. Además de las ventajas económicas que trae este método, se tiene la capacidad de almacenar los datos en tiempo real y la habilidad de transmitirla rápidamente a una estación de control.

Por otra parte, las desventajas o limitaciones que tienen los VANTs están relacionados con la calidad de los sensores, sobre todo a los de bajo costo, ya que son bastante limitados en calidad, peso y dimensión. Es por esto que los drones deben tomar una mayor cantidad de fotografías para cubrir una misma área y obtener datos con una resolución similar [27]. Las fotografías profesionales hechas con drones son tomadas en bloques de múltiples líneas de vuelo con el objetivo de cubrir el área completa. Cada imagen se superpone a la siguiente fotografía en una línea de aproximadamente el 60 % (superposición hacia adelante), mientras que las líneas adyacentes se superponen en un 20 % – 30 % (costado) (Figura 2.6) [3].

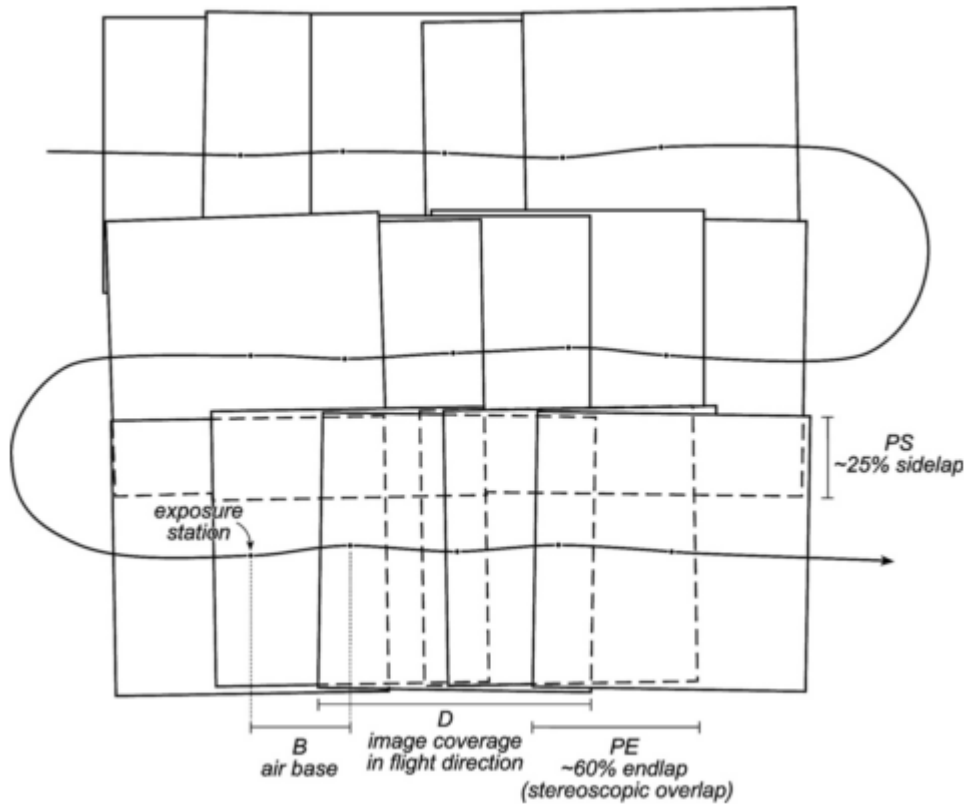


Figura 2.6: Bloque de fotografías aéreas con 60 % de solapamiento, en tres líneas de vuelo con 25 % de solapamiento. Este diseño es ideal para asegurar una cobertura completa y con la redundancia suficiente para aplicaciones posteriores [3].

Por último, es necesario considerar cómo la ley limita el uso de estos vehículos. En el contexto chileno la Dirección General de Aeronáutica Civil (DGAC) reguló el uso de drones en la norma técnica DAN-151, medida provisoria hasta el año 2018, excluyendo a los Drones de uso recreativo. En términos generales fija que para operar un dron se exige ser mayor de edad con autorización previa de la DGAC, la aeronave debe estar registrada y el piloto debe contar con credenciales de la DGAC. Los aparatos deben pesar menos de 6 kilogramos y contar con un paracaídas, y no pueden realizarse vuelos a menos de 2 kilómetros de un aeropuerto o aeródromo, instalaciones de la FACH, áreas de incendio y otros sectores establecidos. Los drones podrán operarse a un máximo de 500 metros del operador, a una altura máxima de 130 metros, de operación manual y siempre mantener contacto visual directo con el vehículo. Finalmente, un piloto durante la operación de un dron no tiene permitido violar los derechos de otras personas en su privacidad y su intimidad [28].

### 2.3. Fotogrametría

La fotogrametría es definida como “el arte, ciencia y tecnología de obtener información de confianza sobre objetos físicos y el ambiente a través del proceso de tomar, medir e interpretar fotografías y patrones de energía electromagnética.” [29]. Esta rama del conocimiento

principalmente realiza medidas precisas de objetos tridimensionales y de características del terreno a partir de fotografías bidimensionales. Las aplicaciones incluyen: medidas de coordenadas; cuantificación de distancias, alturas, áreas y volúmenes; y generación de diversos tipos de mapas topográficos o modelos del terreno. El principio básico tras las medidas de la fotogrametría consiste en la reconstrucción del camino que tomaron los rayos de luz desde el objeto hasta el sensor en el momento de la toma de la fotografía, por tanto, el elemento fundamental es el conocimiento de la geometría característica en una simple imagen [3].

### 2.3.1. Geometría ideal de una fotografía aérea

Una fotografía es el resultado de una proyección central, también conocida como una perspectiva de un punto único. Las distancias del centro óptico del lente de la cámara al sensor en un lado, y al objeto del otro determina la propiedad más básica de una imagen, la escala  $S$ . Para el caso ideal de una fotografía vertical con una perfecta perspectiva central sobre un terreno plano, la Figura 2.7 muestra las relaciones geométricas entre distintas distancias del terreno  $D_i$  y el valor que toman en el plano de la imagen  $d_i$  si es tomada a una altura  $H_g$  y la longitud focal de la cámara es  $f$  [3], variables relacionadas por la Ecuación 2.1 [3].

$$S = \frac{f}{H_g} = \frac{d}{D} \quad (2.1)$$

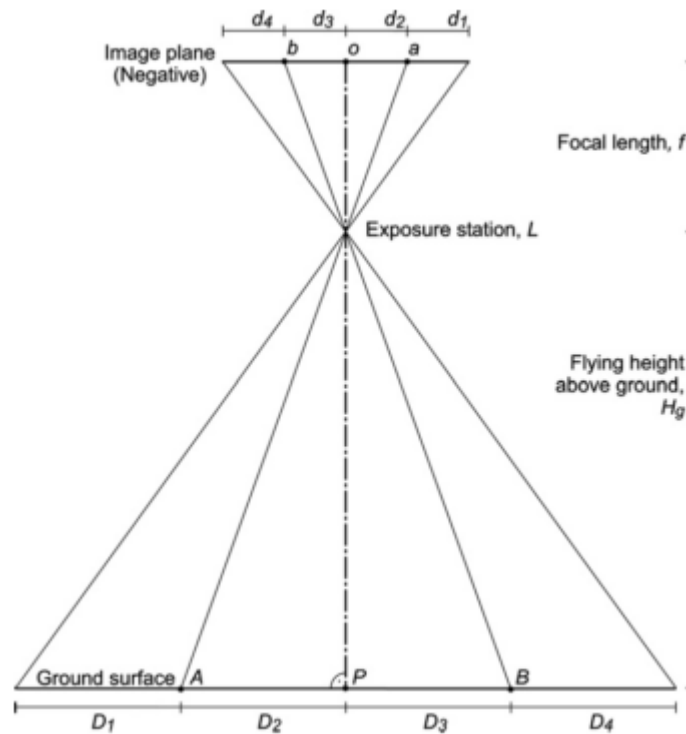


Figura 2.7: Fotografía vertical tomada sobre terreno plano. El eje óptico, que interseca el plano de la imagen en su centro  $o$ , encuentra el suelo en un ángulo recto en el punto principal  $P$ . Obsérvese que las distancias respectivas entre todos los puntos son las mismas en el suelo y en la imagen [3]



### 2.3.2. Geometría real de una fotografía aérea

Situaciones reales a las que se enfrenta la gran mayoría de las fotografías aéreas, en especial SFAP, son bastante diferentes a la descrita en la figura anterior 2.7, ya que el suelo no es completamente plano, es decir, la distancia entre el plano de la imagen y el plano del terreno varía dentro de la imagen. La fotografía no es completamente vertical, ya que el eje óptico no es perpendicular al terreno; y la proyección central es imperfecta porque existen distorsión en los lentes. Estas tres situaciones estropean la semejanza de los triángulos, causando variaciones de escala y distorsiones geométricas de los objetos dentro de la imagen. La Figura 2.8 muestra el efecto que causan diferentes elevaciones en la fotografía vertical, donde todos los puntos que se encuentran en la misma altura presentan la misma escala, los puntos más cercanos tienen una mayor escala y los más lejanos una menor. Al mismo tiempo las posiciones de los puntos en la imagen están desplazadas radialmente hacia fuera del centro para las posiciones altas, y hacia dentro para las más bajas. Este fenómeno llamado desplazamiento por relieve incrementa con la distancia hacia el centro de la imagen, y es inversamente proporcional a la altura de vuelo y al largo focal [3].

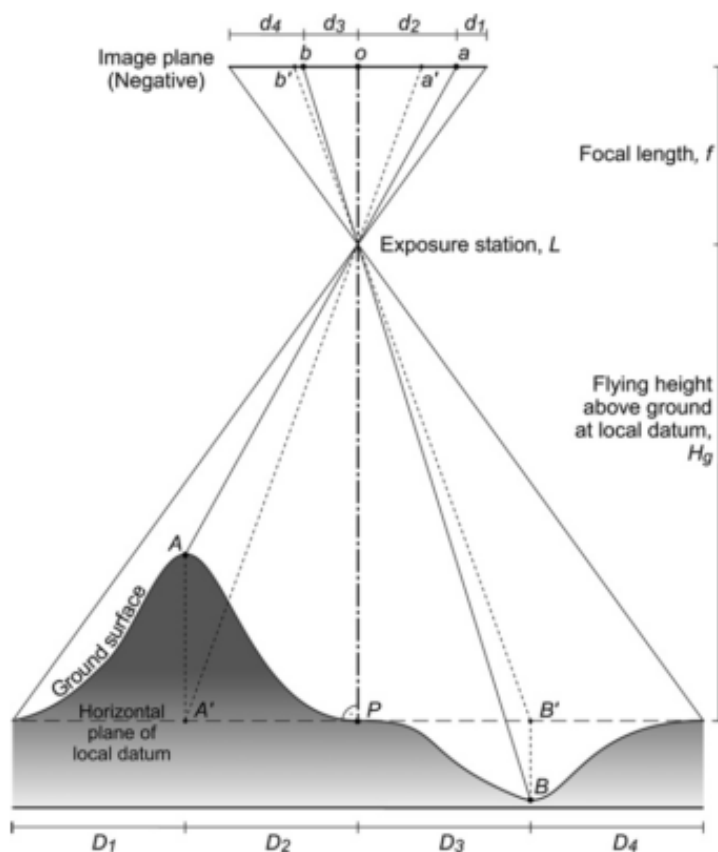


Figura 2.8: Fotografía vertical tomada sobre terreno variable. La elevación del punto principal  $P$  determina el plano horizontal del dato local. Los puntos situados en este plano permanecen sin distorsión, mientras que los puntos arriba o abajo se desplazan radialmente con respecto al centro de la imagen. Observe que las distancias horizontales  $D_1 - D_4$  son las mismas en el espacio de objetos, pero no en la imagen. [3]

La ecuación anterior no sirve para fotografías oblicuas sin un eje óptico vertical, ya que la escala varía según el ángulo de orientación de la inclinación (Figura 2.9). Los principios geométricos recién mencionados son suficientes para realizar mediciones simples a través de fotografías tanto analógicas como digitales. Para el cálculo de coordenadas en tres dimensiones, el camino de los rayos tanto dentro como fuera de la cámara deben ser reconstruidos matemáticamente [3].

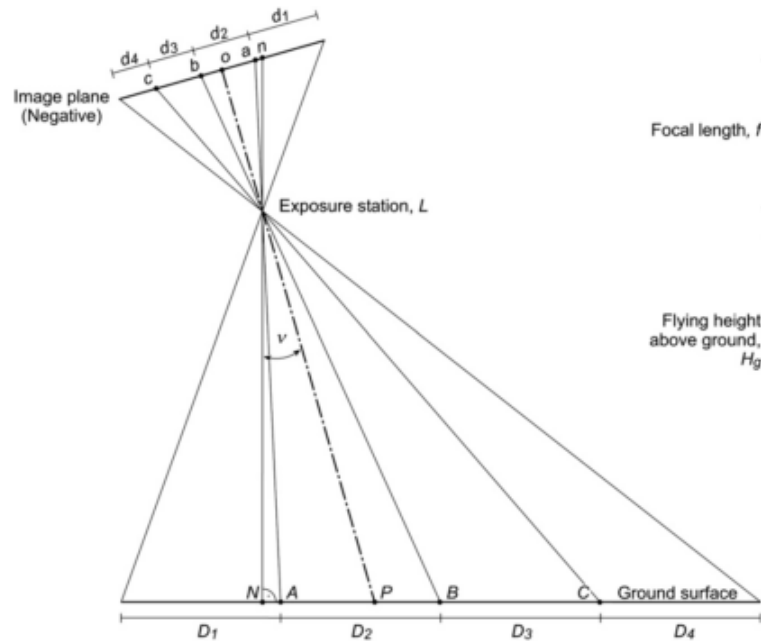


Figura 2.9: Fotografía oblicua tomada sobre un terreno completamente plano. La línea vertical que atraviesa el centro  $L$  en perspectiva interseca el plano de la imagen en el punto nadir fotográfico  $n$  y se encuentra con el suelo en ángulo recto en el punto nadir  $N$ . El ángulo entre la línea nadir y el eje óptico es el ángulo nadir  $v$ . Obsérvese que las distancias  $D_i$  entre todos los puntos en el suelo son iguales, pero las distancias  $d$  correspondientes en la imagen varían continuamente [3]

## 2.4. Procesamiento imágenes

Imágenes SFAP para aplicaciones científicas son tomadas generalmente con la intención de mostrar información sobre un sitio o área, por lo que optimizar la interpretación visual de una imagen con técnicas de mejoramiento suele ser importante. Los defectos de las imágenes pueden ser corregidos parcialmente o completamente con softwares especializados, y aquellos más importantes en este formato de fotografía son [3]:

- Aberraciones radiales: Son defectos circulares centrados en el centro óptico de la imagen, por lo que son importantes de corregir antes de cortar o deformar la imagen. Son distorsiones geométricas específicas del lente (Figura 2.10 A).
- Vignetting: Corresponde a la disminución del brillo de una imagen hacia los bordes de

ésta en comparación a su centro. El grado de vignetting depende de las características del lente, la apertura y la exposición, y no es pues un efecto invariable (Figura 2.10 B).

- Aberraciones cromáticas: Defecto causado por el fenómeno de la dispersión, donde el lente de la cámara falla en enfocar todos los colores en el mismo punto de convergencia (Figura 2.10 C).
- Ruido: Variaciones aleatorias en la información de color de algunos pixeles de una imagen. Puede ocurrir por los circuitos electrónicos de la cámara digital (Figura 2.10 D).

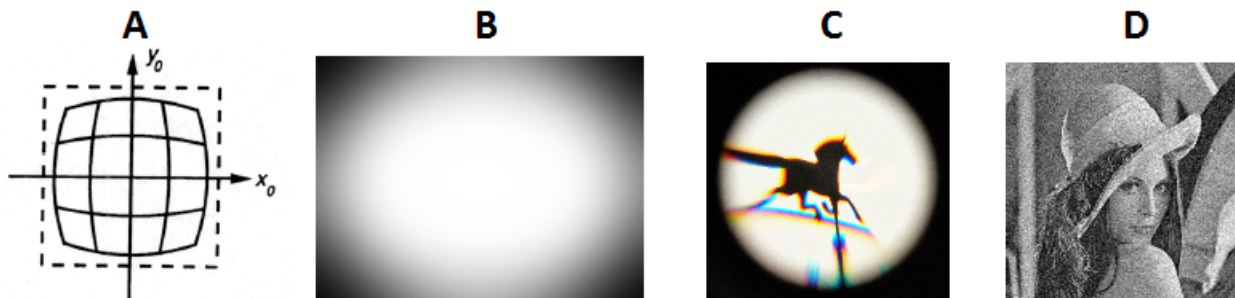


Figura 2.10: Ejemplos de defectos que pueden tener las imágenes. Ejemplo A: Aberraciones radiales. Ejemplo B: Vignetting. Ejemplo C: Aberraciones cromáticas. Ejemplo D: Ruido. [5] [6] [7] [8]

Por otra parte, los objetos presentes en la imagen en sí mismos también pueden ser responsables de defectos en la visualización, especialmente el contraste. El mejoramiento de imágenes puede variar desde ajustes para aumentar la apariencia visual, hasta considerables modificaciones para enfatizar algún tipo de información. A continuación, se explicarán algunos de los métodos de mejoramiento más utilizados en imágenes SFAP [3]:

- Las distorsiones radiales son corregidas con las características específicas del lente y con la distancia focal usando programas dedicados.
- Los efectos del vignetting son algo difícil de corregir automáticamente y su corrección generalmente debe ser juzgada visualmente en cada imagen. Usando deslizadores para la cantidad y el radio de progresión de la iluminación, el brillo de la imagen es modificado hasta que el decaimiento de la luz en las esquinas es compensado.
- La corrección de aberraciones cromáticas involucra transformaciones geométricas de la imagen, donde las aberraciones latitudinales son rectificadas ajustando ligeramente el tamaño de los canales de la imagen modificando radialmente los canales rojos y azul para que calcen con el tamaño del verde.
- El filtrado de imágenes involucra operadores de vecindad, donde el valor de cada pixel en la escena es recalculado en base a los pixeles que lo rodean, con el objetivo de eliminar el ruido de la fotografía.

### 2.4.1. Transformación de imágenes

La transformación de imágenes son operaciones que reexpresan información existente en las mismas, generando un nuevo set de componentes en una imagen representando una descripción de los datos de manera alternativa. Una de las técnicas más utilizadas en el contexto de la agricultura de precisión y la fotografía aérea, corresponde a las relaciones espectrales. Este proceso es el resultado de dividir el valor de un pixel de una banda de la imagen por el valor correspondiente de otra banda de la misma. El efecto que esto tiene es que las variaciones espectrales entre las bandas son mejoradas mientras que las variaciones de brillo son suprimidas. Esto hace que dicho tipo de transformación sea útil para dos propósitos: distinguir de forma más clara entre características con diferencias sutiles de colores, y reducir el efecto de las sombras en escenas donde el sol ilumina. [3].

Una de las aplicaciones más importantes de las relaciones de imágenes corresponde a mejorar las características espectrales de minerales y vegetación. Las superficies vegetativas tienen curvas de reflectancia espectral característica, donde la pendiente pronunciada entre el infrarrojo cercano y las longitudes de onda rojas no aparecen en otros tipos de coberturas, y describirlo con esta técnica usando aquellas bandas espectrales ha sido un método importante para investigar parámetros biofísicos de la vegetación. Los índices de vegetación son diseñados para mostrar, por ejemplo, índice de área foliar, biomasa verde, porcentaje de cobertura, contenido de clorofila, o contenido de proteína mientras se normalizan los efectos de las sombras y variaciones lumínicas. [3].

### 2.4.2. Índices de vegetación para la agricultura de precisión

La respuesta a la luz de la vegetación está directamente relacionada con la existencia de pigmentos, estructura de las hojas, y cantidad de agua, las que varían según la especie, su interacción con el clima, y otros factores. En el espectro visible los pigmentos fotosintéticos son los que condicionan la respuesta espectral de las plantas, entretanto la respuesta de la región del infrarrojo cercano está condicionada por la estructura interna de la hoja. Entender todo esto ha permitido la creación de índices de vegetación para utilizar en la agronomía de precisión [18].

Los índices de vegetación son medidas cuantitativas de la biomasa o vigor vegetal. Usualmente son una combinación de las bandas espectrales, siendo el producto de varios valores espectrales sumados, divididos, o multiplicados en una forma diseñada para producir un único valor. Esto permite estimar y evaluar el estado de salud de la vegetación, en base a la medición de la radiación que las plantas emiten o reflejan [18].

El índice de vegetación más utilizado es el Índice de Vegetación de Diferencia Normalizado (NDVI en inglés), el que se basa en la respuesta que tienen las plantas frente a distintas bandas de frecuencia. La vegetación saludable posee una diferencia notable entre la banda roja y el infrarrojo cercano. Los pigmentos de las plantas absorben gran parte de la luz visible que reciben, en cambio para el infrarrojo cercano las paredes celulares reflejan la mayor parte de la energía cuando poseen mucha agua. Si la planta presenta algún tipo de enfermedad

o deshidratación por otro motivo, la cantidad de agua disminuye por lo que aumentaría la reflexión del color rojo y disminuiría la cantidad reflejada del infrarrojo cercano. El fenómeno explicado permite generar un índice para identificar la vegetación sana de otras, el que se calcula como:

$$NDVI = \frac{NIR - RED}{NIR + RED}, \quad (2.2)$$

donde NIR corresponde al valor reflejado del canal infrarrojo cercano, y RED al valor del canal del rojo visible. Los valores de NDVI están ubicados entre  $-1$  y  $1$  y se considera que por encima de  $0,1$  indican presencia de vegetación, y mientras más cercano a  $1$  mejores son las condiciones de la planta. La Figura 2.11 muestra una imagen procesada por un cálculo NDVI de cada pixel. Las áreas rojas corresponden a sectores ricos en clorofilas, lo que permite discernir de los terrenos cultivados, los posibles problemas que éstos podrían tener, y las áreas sin vegetación. Por ejemplo, el sector del cultivo que posee colores más verdes podría corresponder a un sector cuya tierra no permite una correcta absorción del agua en las plantas o una plaga que está expandiéndose. Esto permite que el agrónomo a cargo del lugar pueda ir a revisar la zona precisa, verificar qué es lo que está sucediendo, y tomar mejores decisiones. Lo que se busca finalmente es aplicar este índice, y otros que vayan al caso, a una imagen de la zona agrícola completa de interés para tener una visión total del escenario de estudio. [18].

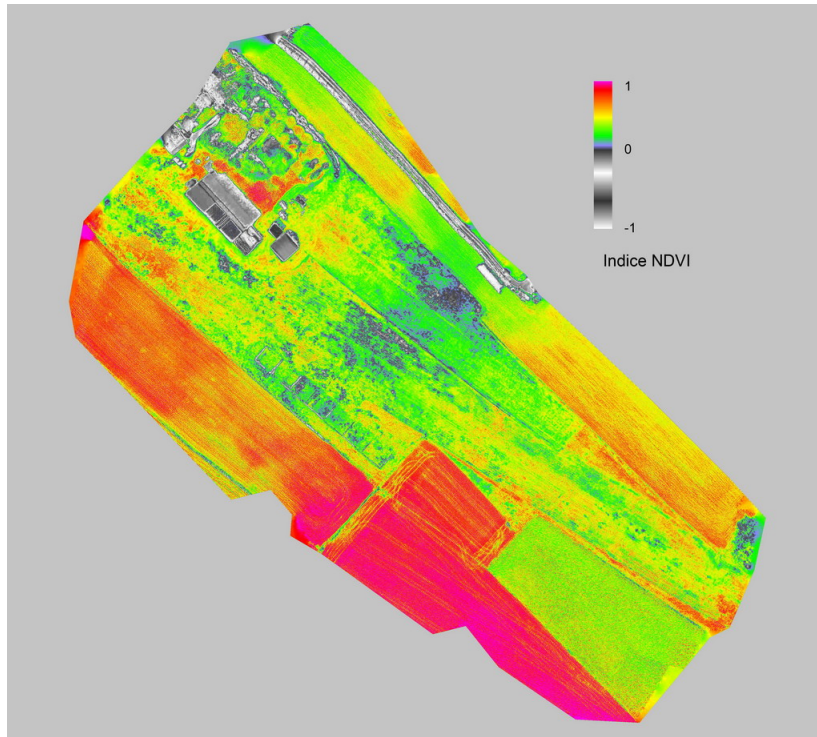


Figura 2.11: Fotografía de un terreno agrícola representado por el índice de vegetación NDVI [9].

## 2.5. Ortomosaicos

Los VANT vuelan a baja altitud, por lo que la cámara de éstos tiene una vista limitada. Para cubrir un área razonablemente grande, la cámara debe tomar cientos de fotografías, motivo por el que para proveer una vista unificada estas fotografías deben ser alineadas y combinadas en una sola. Este proceso de combinar las imágenes es llamado mosaicking, y la imagen resultante es llamada mosaico. Lo que se espera de uno es que esté geométricamente corregido para que pueda ser alineado con un sistema de coordenadas geoespacial. Este flujo de corrección es llamado ortorrectificación de un mosaico llamado mosaico ortorrectificado o simplemente ortomosaico. Se podría llegar a pensar que con la orientación y posición de la cámara por las medidas realizadas vía GPS e IMU, es sencillo generar mosaicos, sin embargo, no es el caso, como se puede ver en la Figura 2.12. Las variaciones de la elevación del terreno, particularmente cerca de un arroyo, barrancos y colinas, además de los errores asociados a las mediciones anteriores, causan malos alineamientos en el mosaico, y también lo hace el lente y cualquiera de las distorsiones mencionadas en secciones anteriores [30].



Figura 2.12: Mosaico rugoso formado a partir de imágenes aéreas usando solamente los datos del GPS y de la IMU [10].

Para generar un ortomosaico existen variadas metodologías, cada una ajustada para la aplicación en que se esté trabajando, la escena tomada y los recursos tecnológicos con que se cuenten. Para este trabajo se desarrollará una metodología que permita una primera aproximación para la generación de ortomosaicos en el contexto de la agricultura, es decir, imágenes tomadas a mediana altura a través de vehículos aéreos no tripulados utilizando cámaras digitales de variadas calidades, y escenas con terrenos frondosos de distinto tipo, donde el mayor interés corresponde a calcular diferentes índices de relación, como los descritos anteriormente, para una zona agrícola. Con todo esto en cuenta, existe una serie de aspectos transversales a todo ortomosaico, que ya se encuentran estudiados y teorizados, los que facilitan la generación de los mismos.

### 2.5.1. SIFT para encontrar puntos claves

El algoritmo Scale-Invariant Feature Transform (SIFT) detecta una gran cantidad de puntos claves en cada imagen aplicando cuatro pasos bien estructurados. Estos puntos detectados en todas las fotografías son útiles al momento de querer buscar regiones solapadas, ya que permite comparar unos pocos píxeles entre imágenes para buscar similitudes. A continuación, se resume el algoritmo SIFT [31]:

1. Detección de puntos extremos: Lo primero que se debe realizar es buscar puntos clave (o de interés) en la imagen completa, es decir, puntos que tengan una definición clara y matemáticamente bien fundada, con una posición bien definida en el espacio de la imagen, que posean estructuras de imagen locales alrededor del punto de interés que sean ricas en contenido de información, ser estables bajo deformaciones locales y globales, y que sean lo suficientemente distintos unos de otros [32].
  - Para ello se generan varios espacios llamados octavas, cada uno creado al reducir sucesivamente el tamaño de la imagen original. Éstas son sometidas progresivamente a un filtro gaussiano, tipo de difuminación que utiliza una función gaussiana  $G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{d^2}{2\sigma^2}}$ , donde  $d$  corresponde a la distancia desde el pixel central de la máscara, para calcular la transformación aplicada a cada pixel [33], obteniendo un subconjunto de imágenes difuminadas para cada octava.
  - Luego de esto, se calcula la resta entre todos los pares consecutivos de imágenes difuminadas (diferencia de gaussianas) dentro de cada octava. Cada pixel de las imágenes nuevas posee 8 vecinos. Si estas imágenes se superponen entre sí, para cada octava, se genera un espacio tridimensional, el cual adiciona a cada pixel 18 vecinos adicionales (9 de la imagen superior y 9 de la imagen inferior), terminando con 26 píxeles vecinos en total. Se seleccionan como candidatos a puntos claves todos los píxeles que sean más grandes o más pequeños que el resto de sus vecinos.
2. Localización precisa de puntos: Para cada candidato a punto de interés, se ajusta un modelo que permite determinar su localización y escala.
  - Para encontrar la posición exacta del candidato a punto clave se realiza una aproximación de segundo orden al rededor del pixel candidato, trabajando a nivel de "subpixel". Encontrando el máximo o mínimo de la aproximación se obtiene la localización del punto clave de manera más precisa.
  - De este conjunto de puntos se eliminan todos aquellos que no correspondan a una esquina, ya que no son útiles como características. Si la magnitud de la intensidad del subpixel en la imagen en que fue encontrado como candidato es menor a cierto umbral, se elimina. Por otra parte, para eliminar bordes y regiones planas se hace un cálculo de gradientes perpendiculares entre sí, sobre el punto clave. Si ambos gradientes son bajos, corresponde a una región plana; y si uno es grande y el otro pequeño, es un borde.
3. Asignación de la orientación: A cada punto característico se le asigna una o varias orientaciones según las orientaciones y magnitudes del gradiente local.
  - Hasta este punto ya se conoce la escala a la que fue obtenido cada punto de interés. Ahora se procede a encontrar direcciones y magnitudes de gradientes alrededor de cada uno. Cuando esto se realiza para los puntos vecinos se crea un histograma de

orientaciones, seleccionando como dirección del punto clave aquel más alto. Vale acotar que si hay uno cuyo valor es superior al 80 %, es considerado como un nuevo punto de interés con misma escala y posición, pero diferente dirección.

4. Descripción del punto característico: Alrededor de cada punto característico se representa su región de tal forma, que sea robusta a cambios significativos en la iluminación y a pequeñas distorsiones en la forma.
  - Finalmente se debe representar el punto clave en un descriptor, vector que represente la información de éste, y que sea capaz de diferenciarse de otros. Para ello se toman los  $16 \times 16$  píxeles al rededor del punto clave, los que se dividen en 16 ventanas. Dentro de cada una se calculan las orientaciones y magnitudes de los gradientes, y se genera un nuevo histograma, como el anterior, pero los pesos varían según la distancia a la que se encuentre del punto clave.
  - Una vez hecho esto se concatenan los histogramas, obteniendo un vector que al normalizarlo se transforma en el descriptor del punto.

### 2.5.2. Matriz de Homografía para superponer imágenes

Una homografía corresponde a cualquier transformación proyectiva que genere una correspondencia entre dos figuras geométricas planas (fotografías), tal que a cada punto de ellas le corresponda un punto de la otra. Para definir cómo se realiza esta relación geométrica hay que representar los puntos de una imagen de la siguiente manera: Un punto  $(x, y)$  en una imagen puede representarse por un vector  $P = (p_1, p_2, p_3)$ , donde  $x = \frac{p_1}{p_3}$  e  $y = \frac{p_2}{p_3}$ , lo que se llama representación homogénea de un punto. La homografía entre dos puntos equivalentes  $P'$  y  $P$  de dos figuras se calcula como  $P' = HP$ , donde  $P'$  corresponde a un punto de una imagen emparejado con el punto  $P$  de la otra, y  $H$  corresponde a la matriz de homografía. Si se realiza esta transformación con la matriz adecuada sobre dos imágenes completas que compartan información, se puede lograr superponer las fotografías para generar el ortomosaico. Esto significa que para calcular la homografía que mapea un punto  $P$  a su correspondiente  $P'$  en otra imagen, basta con calcular la matriz de homografía de  $3 \times 3$  (Figura 2.13) [34] [35].

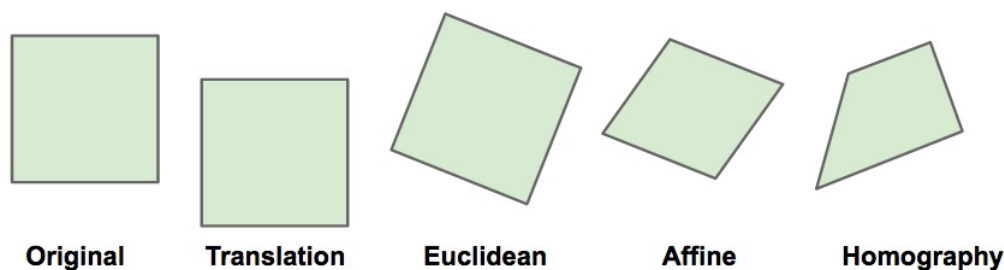


Figura 2.13: Una homografía puede tener distintos tipos de transformaciones visuales según los grados de libertad que esta tenga. El primer caso es la original, luego la traslación con 2 grados de libertad, la transformación euclidiana para 3 grados de libertad, transformación afín para 6 grados de libertad y homografía para 8 [11].



La matriz  $H$  produce proyecciones invariables en escala, demostración que se hará a continuación. Para ello se debe probar que la coordenada  $(x, y)$  obtenida por el vector  $P'$  generado por la proyección  $P'_a = HP$  es el mismo que el generado por  $P'_b = \lambda HP$  [36]:

$$P'_a = \begin{pmatrix} p'_{a1} \\ p'_{a2} \\ p'_{a3} \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \quad (2.3)$$

$$P'_a = \begin{pmatrix} H_{11}p_1 + H_{12}p_2 + H_{13}p_3 \\ H_{21}p_1 + H_{22}p_2 + H_{23}p_3 \\ H_{31}p_1 + H_{32}p_2 + H_{33}p_3 \end{pmatrix} \quad (2.4)$$

$$x'_a = \frac{p'_{a1}}{p'_{a3}} = \frac{H_{11}p_1 + H_{12}p_2 + H_{13}p_3}{H_{31}p_1 + H_{32}p_2 + H_{33}p_3} \quad (2.5)$$

$$y'_a = \frac{p'_{a2}}{p'_{a3}} = \frac{H_{21}p_1 + H_{22}p_2 + H_{23}p_3}{H_{31}p_1 + H_{32}p_2 + H_{33}p_3} \quad (2.6)$$

Luego, se realiza el mismo procedimiento para el punto  $P'_b$  para verificar la igualdad:

$$P'_b = \begin{pmatrix} p'_{b1} \\ p'_{b2} \\ p'_{b3} \end{pmatrix} = \lambda \begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \quad (2.7)$$

$$P'_b = \begin{pmatrix} \lambda H_{11}p_1 + \lambda H_{12}p_2 + \lambda H_{13}p_3 \\ \lambda H_{21}p_1 + \lambda H_{22}p_2 + \lambda H_{23}p_3 \\ \lambda H_{31}p_1 + \lambda H_{32}p_2 + \lambda H_{33}p_3 \end{pmatrix} \quad (2.8)$$

$$x'_b = \frac{p'_{b1}}{p'_{b3}} = \frac{\lambda H_{11}p_1 + \lambda H_{12}p_2 + \lambda H_{13}p_3}{\lambda H_{31}p_1 + \lambda H_{32}p_2 + \lambda H_{33}p_3} = \frac{H_{11}p_1 + H_{12}p_2 + H_{13}p_3}{H_{31}p_1 + H_{32}p_2 + H_{33}p_3} = x'_a \quad (2.9)$$

$$y'_b = \frac{p'_{b2}}{p'_{b3}} = \frac{\lambda H_{21}p_1 + \lambda H_{22}p_2 + \lambda H_{23}p_3}{\lambda H_{31}p_1 + \lambda H_{32}p_2 + \lambda H_{33}p_3} = \frac{H_{21}p_1 + H_{22}p_2 + H_{23}p_3}{H_{31}p_1 + H_{32}p_2 + H_{33}p_3} = y'_a \quad (2.10)$$

Demostrando finalmente que la matriz  $H$  genera la misma proyección independiente de la escala. Esto permite que se pueda definir esta matriz con solo 8 grados de libertad [37]:

$$H = \begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & 1 \end{pmatrix} \quad (2.11)$$

Ya definida la matriz de homografía y sus propiedades más importantes, junto al uso que puede tener en la generación del ortomosaico, es necesario entender cómo se puede estimar. Considerando que la  $H$  posee 8 grados de libertad, son 8 las incógnitas a resolver por pares de

imágenes. Con todo esto en consideración, se procede a desarrollar las ecuaciones según pares de emparejamientos  $(X_i, Y_i)$  y  $(X'_i, Y'_i)$  a partir de la relación  $P' = HP$ , donde  $P'_i = (x'_i, y'_i, w'_i)$ ,  $P_i = (x_i, y_i, w_i)$ ,  $X'_i = \frac{x'_i}{w'_i}$ ,  $Y'_i = \frac{y'_i}{w'_i}$ ,  $X_i = \frac{x_i}{w_i}$ , e  $Y_i = \frac{y_i}{w_i}$ . Con esto se representan los vectores  $P$  y  $P'$  de la siguiente manera [37]:

$$P'_i = \begin{pmatrix} X'_i \cdot w'_i \\ Y'_i \cdot w'_i \\ w'_i \end{pmatrix} \quad (2.12)$$

$$P_i = \begin{pmatrix} X_i \cdot w_i \\ Y_i \cdot w_i \\ w_i \end{pmatrix} \quad (2.13)$$

Reemplazando en la relación  $P'_i = HP_i$  se tiene:

$$\begin{pmatrix} X'_i \cdot w'_i \\ Y'_i \cdot w'_i \\ w'_i \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & 1 \end{pmatrix} \begin{pmatrix} X_i \cdot w_i \\ Y_i \cdot w_i \\ w_i \end{pmatrix} \quad (2.14)$$

Luego, desarrollando la multiplicación:

$$\begin{pmatrix} X'_i \cdot w'_i \\ Y'_i \cdot w'_i \\ w'_i \end{pmatrix} = \begin{pmatrix} X_i \cdot w_i \cdot H_{11} + Y_i \cdot w_i \cdot H_{12} + w_i \cdot H_{13} \\ X_i \cdot w_i \cdot H_{21} + Y_i \cdot w_i \cdot H_{22} + w_i \cdot H_{23} \\ X_i \cdot w_i \cdot H_{31} + Y_i \cdot w_i \cdot H_{32} + w_i \end{pmatrix} \quad (2.15)$$

Finalmente:

$$\begin{cases} X'_i = \frac{X'_i \cdot w'_i}{w'_i} = \frac{X_i \cdot w_i \cdot H_{11} + Y_i \cdot w_i \cdot H_{12} + w_i \cdot H_{13}}{X_i \cdot w_i \cdot H_{31} + Y_i \cdot w_i \cdot H_{32} + w_i} = \frac{X_i \cdot H_{11} + Y_i \cdot H_{12} + H_{13}}{X_i \cdot H_{31} + Y_i \cdot H_{32} + 1} \\ Y'_i = \frac{Y'_i \cdot w'_i}{w'_i} = \frac{X_i \cdot w_i \cdot H_{21} + Y_i \cdot w_i \cdot H_{22} + w_i \cdot H_{23}}{X_i \cdot w_i \cdot H_{31} + Y_i \cdot w_i \cdot H_{32} + w_i} = \frac{X_i \cdot H_{21} + Y_i \cdot H_{22} + H_{23}}{X_i \cdot H_{31} + Y_i \cdot H_{32} + 1} \end{cases} \quad (2.16)$$

$$\Rightarrow \begin{cases} X'_i \cdot (X_i \cdot H_{31} + Y_i \cdot H_{32} + 1) = X_i \cdot H_{11} + Y_i \cdot H_{12} + H_{13} \\ Y'_i \cdot (X_i \cdot H_{31} + Y_i \cdot H_{32} + 1) = X_i \cdot H_{21} + Y_i \cdot H_{22} + H_{23} \end{cases} \quad (2.17)$$

$$\Rightarrow \begin{cases} X_i \cdot H_{11} + Y_i \cdot H_{12} + H_{13} - X'_i X_i \cdot H_{31} - X'_i Y_i \cdot H_{32} = X'_i \\ X_i \cdot H_{21} + Y_i \cdot H_{22} + H_{23} - Y'_i X_i \cdot H_{31} - Y'_i Y_i \cdot H_{32} = Y'_i \end{cases} \quad (2.18)$$

Obteniendo dos ecuaciones lineales con las coordenadas originales de las imágenes. Si esto se aplica para cuatro emparejamientos, se tendrá un conjunto de 8 ecuaciones, lo que permite resolver el sistema y obtener así la matriz de homografía  $H$  que proyecta una imagen en el plano de la otra. Con esto se puede armar el siguiente sistema de ecuaciones lineales de forma matricial, el que se puede resolver fácilmente [38]:

$$\begin{pmatrix} X_1 & Y_1 & 1 & 0 & 0 & 0 & -X_1 X'_1 & -X'_1 Y_1 \\ 0 & 0 & 0 & X_1 & Y_1 & 1 & -X_1 Y'_1 & -Y_1 Y'_1 \\ X_2 & Y_2 & 1 & 0 & 0 & 0 & -X_2 X'_2 & -X'_2 Y_2 \\ 0 & 0 & 0 & X_2 & Y_2 & 1 & -X_2 Y'_2 & -Y_2 Y'_2 \\ X_3 & Y_3 & 1 & 0 & 0 & 0 & -X_3 X'_3 & -X'_3 Y_3 \\ 0 & 0 & 0 & X_3 & Y_3 & 1 & -X_3 Y'_3 & -Y_3 Y'_3 \\ X_4 & Y_4 & 1 & 0 & 0 & 0 & -X_4 X'_4 & -X'_4 Y_4 \\ 0 & 0 & 0 & X_4 & Y_4 & 1 & -X_4 Y'_4 & -Y_4 Y'_4 \end{pmatrix} \begin{pmatrix} H_{11} \\ H_{12} \\ H_{13} \\ H_{21} \\ H_{22} \\ H_{23} \\ H_{31} \\ H_{32} \end{pmatrix} = \begin{pmatrix} X'_1 \\ Y'_1 \\ X'_2 \\ Y'_2 \\ X'_3 \\ Y'_3 \\ X'_4 \\ Y'_4 \end{pmatrix} \quad (2.19)$$

### 2.5.3. RANSAC para la eliminación de emparejamientos erróneos

El algoritmo RANSAC (RANDOM SAMPLE CONSENSUS) es un enfoque de estimación de parámetros generales diseñado para encontrar y eliminar valores atípicos en una serie de datos explicados por un modelo, para el tema de este trabajo, la matriz de homografía. RANSAC es una técnica que genera candidatos de solución utilizando la menor cantidad de datos (emparejamientos) posibles, los necesarios para estimar el modelo que subyace en el problema (Figura 2.14). El algoritmo tiene la siguiente estructura [39]:

1. Se seleccionan aleatoriamente los puntos requeridos para determinar los parámetros del modelo.
2. Se obtiene el modelo con los puntos seleccionados.
3. Se calcula cuántos puntos del conjunto completo calzan bien en el modelo con cierto nivel de tolerancia.
4. Si la fracción de inliers sobre el total de puntos supera cierto umbral, se reestima el modelo con estos puntos y se eliminan los outliers. De lo contrario, se repite desde el paso 1 al 4 un número determinado de veces, quedándose con el mejor modelo.

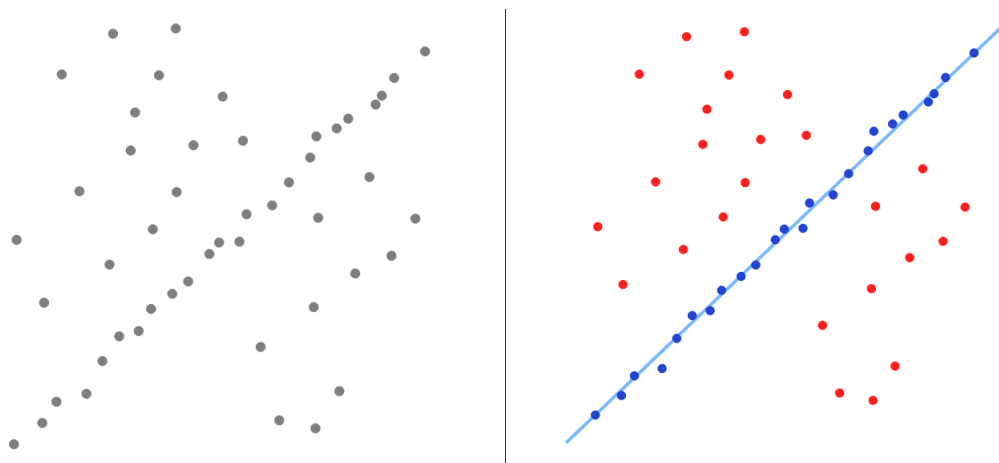


Figura 2.14: A la izquierda se tiene un conjunto de datos con distintos valores atípicos. En la derecha el algoritmo RANSAC encuentra un modelo que explica la mayoría de los datos y permite identificar cuáles son los valores atípicos [12].

### 2.5.4. Compensación de ganancia para ajustar intensidades de imágenes solapadas

Una compensación de ganancia consiste en adecuar las intensidades de las imágenes multiplicando cada uno de sus píxeles por un número real, para llevarlos a un valor deseado. Expresando lo anterior de manera matemática, para una imagen  $I$  cada uno de sus píxeles se multiplica por la ganancia deseada  $g$ , obteniendo finalmente la matriz  $g \cdot I$ . Vale acotar que todo píxel que exceda el valor máximo que puede tomar, cifra establecida según el número de bits por el que está representado, se hará igual a este.

Las principales fuentes que producen las diferencias de intensidad entre fotografías, a pesar de referirse a la misma escena, se pueden resumir en [40]:

- Los ajustes automáticos de la cámara como el control de ganancia automático, balance de blancos o tiempo de exposición;
- Los cambios de iluminación debido a las variaciones de los niveles de luminosidad ambiente, o debido al movimiento relativo entre la escena y la fuente de luz.

Por estos fenómenos, las imágenes con regiones solapadas poseen diferencias de niveles de intensidad a pesar de referirse al mismo punto. Si se desea que las intensidades de pixeles superpuestos sean iguales, para no tener diferencias al momento de superponerlos, es necesario realizar un ajuste de ganancias en cada uno.

Con ese objetivo en mente, el trabajo [14] define una función error establecida por la diferencia de intensidades entre los pixeles de la región compartida de las imágenes superpuestas. Como se busca minimizar esta diferencia, cada intensidad es multiplicada por una ganancia variable, que es la que se busca encontrar para minimizar la expresión. En lugar de comparar la ganancia de cada pixel, se compara la intensidad promedio de cada región solapada entre ambas imágenes, con el objetivo de reducir el costo computacional, y atenuar el efecto que podría otorgar algún outlier sobre los resultados. Este problema trae como formulación matemática:

$$error = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M |R(i, j)| \cdot (g_i \bar{I}_{ij} - g_j \bar{I}_{ji})^2 \quad (2.20)$$

Donde  $R(i, j)$  es la región solapada entre las imágenes  $I_i$  e  $I_j$ ,  $|R(i, j)|$  corresponde a la cantidad de pixeles que se encuentran en la región  $R(i, j)$ ,  $\bar{I}_{ij} = \frac{\sum_{u_i \in R(i, j)} I_i(u_i)}{|R(i, j)|}$  es el promedio de intensidades que existe en la región  $R(i, j)$  de la imagen  $I_i$ , y  $g_i$  es la ganancia de ajuste para la imagen  $i$ -ésima [14]

Es necesario notar que si se minimiza esa función, el resultado  $g = 0$  sería el óptimo, lo que no tiene sentido para lo que se busca. Por este motivo se agrega el término  $(1 - g_i)^2$  para impedir que  $g_i$  sea cero. Además, se le debe agregar un peso  $\lambda_1$  y  $\lambda_2$  a cada término según la importancia relativa a cada objetivo, quedando la función de la siguiente forma [14]:

$$error = \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M |R(i, j)| \cdot \left( \lambda_1 (g_i \bar{I}_{ij} - g_j \bar{I}_{ji})^2 + \lambda_2 (1 - g_i)^2 \right) \quad (2.21)$$

Los valores propuestos para obtener buenos resultados son  $\lambda_1 = 0,01$  y  $\lambda_2 = 100$  según [14]. Con todo esto, se tiene una función cuadrática que se resuelve derivandola en función de

todas las incógnitas e igualándola a cero:

$$\begin{aligned}
\frac{\partial e}{\partial g_i} &= \sum_{k=1, k \neq i}^M |R(i, k)| [\lambda_1 (g_i \bar{I}_{ik} - g_k \bar{I}_{ki})^2 + \lambda_2 (1 - g_i)^2] \\
&+ \frac{1}{2} \sum_{k=1, k \neq i}^M |R(k, i)| [\lambda_1 (g_k \bar{I}_{ki} - g_i \bar{I}_{ik}) + \lambda_2 (1 - g_k)^2] \\
&+ \frac{1}{2} |R(i, i)| (\lambda_2 (1 - g_i)^2)
\end{aligned} \tag{2.22}$$

$$\frac{\partial e}{\partial g_i} = g_i \left[ \sum_{k=1}^M |R(i, k)| \cdot (2\lambda_1 \bar{I}_{ik}^2 + \lambda_2) \right] - \left[ \sum_{k=1, k \neq i}^M g_k \cdot 2|R(k, i)| \lambda_1 \bar{I}_{ki} \bar{I}_{ik} \right] - \left[ \sum_{k=1}^M |R(i, k)| \cdot \lambda_2 \right] \tag{2.23}$$

$$\Rightarrow g_i \alpha_i + \sum_{k=1}^M g_k \beta_i k - \gamma_i = 0 \tag{2.24}$$

Con el cálculo recién realizado se puede expresar el sistema de ecuaciones de manera matricial:

$$\begin{pmatrix} \alpha_1 & \beta_{12} & \dots & \beta_{1(M-1)} & \beta_{1M} \\ \beta_{21} & \alpha_2 & \dots & \beta_{2(M-1)} & \beta_{2M} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta_{(M-1)1} & \beta_{(M-1)2} & \dots & (\alpha_{(M-1)}) & \beta_{(M-1)M} \\ \beta_{M1} & \beta_{M2} & \dots & \beta_{(M-1)M} & (\alpha_M) \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_{(M-1)} \\ g_M \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{(M-1)} \\ \gamma_M \end{pmatrix} \tag{2.25}$$

Se puede ver que el sistema de ecuaciones es de la forma  $Ax = b$ , por lo que se resuelve de la misma forma que el sistema de la sección anterior.

### 2.5.5. Mezcla de imágenes en múltiples bandas para reducir efectos no deseados

La mezcla de imágenes consiste en realizar cambios de intensidades en cada una de éstas para que visualmente parezcan ser una sola. Utilizando dicha técnica es que se puede mejorar la calidad de una imagen compuesta por varias fotografías superpuestas, logrando que los bordes de cada una dejen de ser perceptible (Figura 2.15) [41]. El trabajo [14] propone una técnica llamada mezcla de imágenes en múltiples bandas, enfoque que reduce los errores para las componentes de alta frecuencia, generando mejores resultados. La idea de esto es combinar las componentes de baja frecuencia en rangos espaciales largos, y las de alta frecuencia sobre rangos cortos.

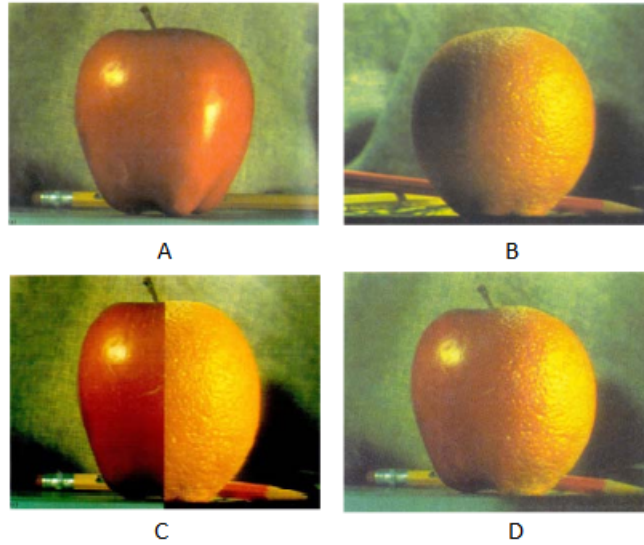


Figura 2.15: Ejemplo de mezcla de imágenes. Figura A y B, corresponden a las imágenes a mezclar. Figura C es el caso en que se pone una mitad sobre la otra sin técnica alguna de mezcla. Figura D es el caso en que se ocupa una técnica de mezcla, haciendo que los bordes de ambas dejen de percibirse [13].

Se empieza asignando un peso para cada pixel de las imágenes de manera lineal, siendo 1 en el centro y 0 en los bordes, generando una matriz de pesos  $W(x, y)$  para cada imagen. Esto se modela así, ya que dentro de los efectos que se desean arreglar, los pixeles centrales guardan de mejor forma la información de la escena capturada. Luego en el ortomosaico ya armado se genera una máscara de pesos para cada imagen obteniéndose así [14]:

$$W_{mask}^i(x, y) = \begin{cases} 1, & \text{si } W^i(x, y) = \arg \max_j W_{mask}^j(x, y) \\ 0, & \text{si no} \end{cases} \quad (2.26)$$

donde  $W^i(x, y)$  corresponde a la matriz de pesos de la imagen  $i$ -ésima. La expresión anterior que decir que  $W_{mask}^i$  es 1 para los pixeles  $(x, y)$  en que los pesos de la imagen  $i$ -ésima sean los más altos, y 0 donde otra imagen sea que posea el mayor [14].

Después de tener todas las máscaras  $W_{mask}(x, y)$ , se aplica un filtro gaussiano a la imagen completa, el que se irá variando a un nuevo rango de frecuencias espaciales. Este procedimiento de filtrado se realiza 5 veces, con filtros de escalas distintas, obteniendo así 5 bandas diferentes llamadas pesos de combinación para cada banda. Se parte formando una versión pasa altos de la imagen:

$$B_\sigma^i(x, y) = I^i(x, y) - I_\sigma^i(x, y)$$

$$I_\sigma^i(x, y) = I^i(x, y) * g_\sigma(x, y)$$

donde  $g_\sigma(x, y)$  corresponde a un filtro gaussiano de desviación estándar  $\sigma$ , y  $B_\sigma(x, y)$  representa las frecuencias espaciales en el rango de longitudes de onda  $[0, \sigma)$ . Se mezcla esta banda entre las imágenes, ponderada por sus pesos correspondientes, obtenidos al pasar el mismo filtro gaussiano sobre su máscara:

$$W_\sigma^i(x, y) = W_{mask}^i(x, y) * g_\sigma(x, y)$$

donde  $W_{\sigma}^i$  son los pesos de combinación para la banda de longitud de onda  $\lambda \in [0, \sigma)$ . Las subsiguientes bandas de frecuencia se combinan usando filtros pasa bajos y volviendo a difuminar las imágenes:

$$\begin{aligned} B_{(k+1)\sigma}^i &= I_{k\sigma}^i - I_{(k+1)\sigma}^i \\ I_{(k+1)\sigma}^i &= I_{k\sigma}^i * g_{\sigma'} \\ W_{(k+1)\sigma}^i &= W_{k\sigma}^i * g_{\sigma'} \end{aligned}$$

donde el filtro gaussiano es calculado con una desviación estándar variable  $\sigma' = \sqrt{(2k+1)\sigma}$ , de tal forma que las bandas subsiguientes tengan el mismo rango de longitudes de onda [14].

Finalmente para cada banda, las imágenes sobrepuestas son combinadas linealmente usando el correspondiente peso de combinación:

$$I_{k\sigma}^{multi}(x, y) = \frac{\sum_{i=1}^M B_{k\sigma}^i \cdot W_{k\sigma}^i(x, y)}{\sum_{i=1}^M W_{k\sigma}^i(x, y)}$$

Esto causa que la banda de alta frecuencia (pequeño  $k\sigma$ ) son mezcladas sobre rangos cortos, mientras que las bandas de baja frecuencia (largo  $k\sigma$ ) son mezcladas sobre rangos largos [14].

# Capítulo 3

## Metodología

En el siguiente capítulo se explicará la metodología diseñada e implementada para generar ortomosaicos de manera automática. El algoritmo generado contempla la creación de ortomosaicos para imágenes tomadas en terreno agrícola en posiciones verticales. Trabajar en este formato permite tener imágenes relativamente perpendiculares a la cámara, lo que facilita la ortorrectificación del ortomosaico.

En la metodología se comienza trabajando con las imágenes en bruto, de manera que utilizar la información geoespacial más la inclinación de la cámara, sea útil para determinar las regiones de la fotografía que presentan información de la cara superior de las plantas, y además determinar cuáles son las fotografías que deberían poseer solapamiento. Luego se buscan sobre estas imágenes con mayor certeza, cuáles son las que comparten información para poder modificar la imagen ya sea rotándola, modificando su escala, o estirándola, para que al superponerlas se empiece a crear el ortomosaico. Después se realiza una adecuación de las imágenes para que los bordes de éstas al superponerse sobre otra no se noten y se vea una fotografía más homogénea, obteniendo así, el ortomosaico. Finalmente se comentan ciertas métricas para medir el desempeño del algoritmo diseñado, lo que será evaluado y analizado en el Capítulo 4.

### 3.1. Geometría de la cámara en la agricultura

En el capítulo anterior se mencionó que la geometría de una fotografía en una situación no ideal puede producir una serie de complicaciones. La primera es que éstas nunca serán completamente tomadas de manera perpendicular a la superficie, estando la cámara siempre rotada levemente hacia un costado tanto en cabeceo (pitch), alabeo (roll), y guiñada (yaw). Además, se debe considerar que todos los cuerpos al no encontrarse en el punto perpendicular de la cámara se alcanzan a fotografiar partes de las caras laterales de los mismos. Por todos estos motivos es que se analiza matemáticamente el impacto que estas variables tienen sobre la fotografía tomada en la Figura 3.1.



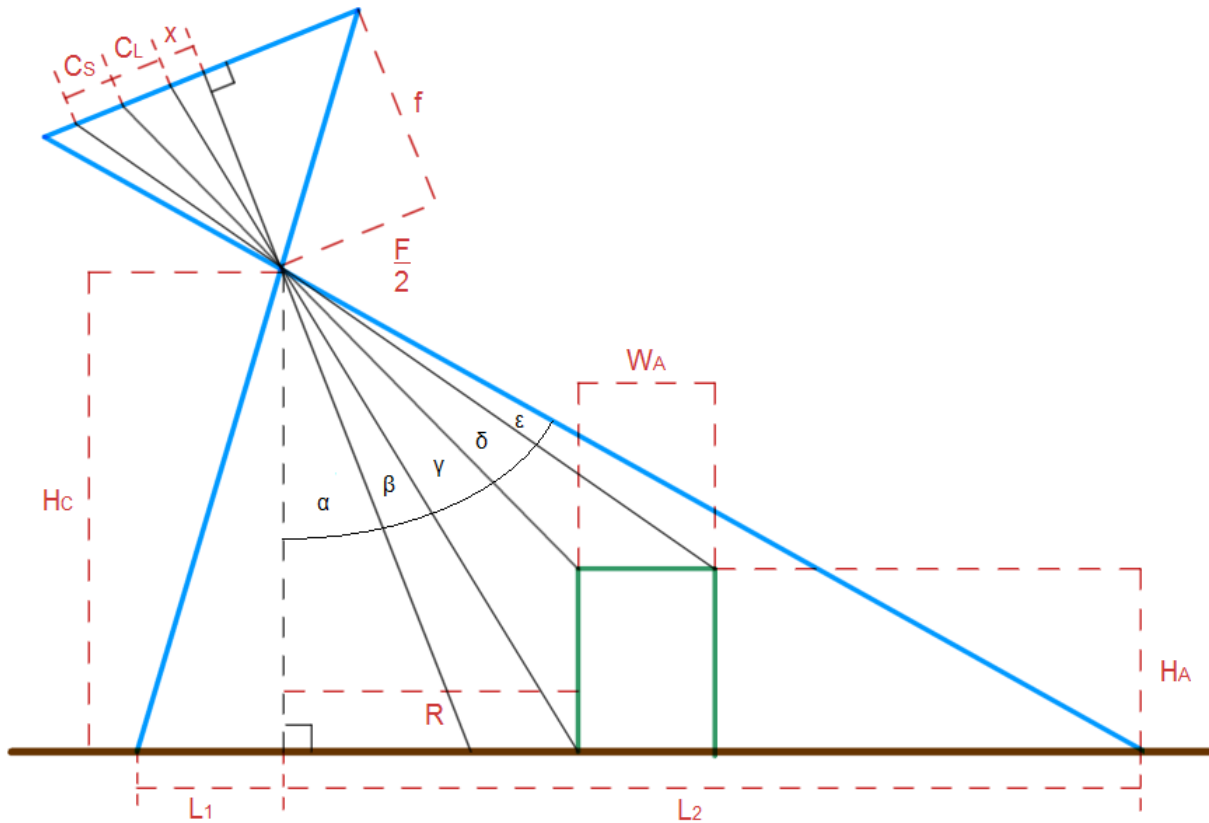


Figura 3.1: Representación de la geometría de una cámara al tomar una fotografía sobre terreno agrícola, donde el objeto verde representa un árbol captado completamente por la misma.

Para facilitar el análisis se divide la rotación en dos ejes relativos a la orientación de la rotación, donde la cámara se encuentra ladeada un ángulo  $\alpha_x$  para el alabeo (eje  $x$ ), y un ángulo  $\alpha_y$  para el cabeceo (eje  $y$ ). Además, el aparato se encuentra a una altura  $H_C$ , su sensor posee un largo  $F_x$ , un ancho  $F_y$ , y una longitud focal  $f$ . Por otra parte, la planta fotografiada se encuentra a una distancia  $R$  desde el punto perpendicular, cuya altura es  $H_A$  y ancho  $W_A$ , objeto que al ser fotografiado queda captado en  $C_L$  y  $C_S$ , segmentos de pixeles que corresponden a la sección de la cara lateral y a la cara superior del árbol respectivamente, captada por la cámara. Por otra parte,  $L_{1x}$ ,  $L_{2x}$ , y  $L_{1y}$ ,  $L_{2y}$  corresponden a las amplitudes completas que pueden ser fotografiadas por la cámara según el eje que se esté considerando, debiendo hacerse distinción luego entre lo que se capta en cada uno según corresponda. Los demás valores son variables auxiliares para el cálculo de los aspectos de interés en la foto resultante, es decir, el cálculo de la cobertura total de la imagen sobre el terreno; y calcular el cociente entre la cara lateral fotografiada y la cara superior de una planta.

El objetivo de estos cálculos es comprender hasta qué pixel de las imágenes los cuerpos pueden ser considerados perpendiculares a la fotografía, exigiendo que el cociente entre ambas caras captadas sea inferior a un valor establecido, lo que llamaremos criterio de perpendicularidad. Esto permite recortar la imagen y trabajar solo con los pixeles útiles, o sea, aquellos que cumplen el criterio mencionado. Además, este procedimiento permite comprender cuales imágenes adyacentes tomadas por la cámara del dron podrían llegar a compartir información

entre si para las etapas posteriores. Como se pudo ver, los cálculos para el eje  $x$  e  $y$  son análogos, motivo por el que se realizan una sola vez, considerando variables generales como  $\alpha$ ,  $F$ ,  $L_1$  y  $L_2$ , las que al momento de la implementación deben ser diferenciadas respectivamente. Con todo esto aclarado, se comienza calculando  $L_2$  aprovechando el triángulo rectángulo formado por el centro de la cámara:

$$\left. \begin{aligned} \tan(\beta + \gamma + \delta + \varepsilon) &= \frac{F}{2f} \\ \tan(\alpha + \beta + \gamma + \delta + \varepsilon) &= \frac{L_2}{H_C} \end{aligned} \right\} L_2 = H_C \cdot \tan \left( \tan^{-1} \left( \frac{F}{2f} \right) + \alpha \right) \quad (3.1)$$

Luego, se obtiene  $L_1$  utilizando semejanza de triángulos y lo escrito en la ecuación anterior:

$$\left. \begin{aligned} \tan(\beta + \gamma + \delta + \varepsilon - \alpha) &= \frac{L_1}{H_C} \\ \beta + \gamma + \delta + \varepsilon &= \tan^{-1} \left( \frac{F}{2f} \right) \end{aligned} \right\} L_1 = H_C \cdot \tan \left( \tan^{-1} \left( \frac{F}{2f} \right) - \alpha \right) \quad (3.2)$$

Ahora se calculan los ángulos  $\beta$ ,  $\gamma$ , y  $\delta$  utilizando las mismas propiedades matemáticas que antes, ya que es necesario para desarrollos posteriores:

- $\tan(\alpha + \beta) = \frac{R}{H_C} \Rightarrow \beta = \tan^{-1} \left( \frac{R}{H_C} \right) - \alpha$
- $\tan(\alpha + \beta + \gamma) = \frac{R}{H_C - H_A} \Rightarrow \gamma = \tan^{-1} \left( \frac{R}{H_C - H_A} \right) - \alpha - \beta$
- $\tan(\alpha + \beta + \gamma + \delta) = \frac{R + W_A}{H_C - H_A} \Rightarrow \delta = \tan^{-1} \left( \frac{R + W_A}{H_C - H_A} \right) - \alpha - \beta - \gamma$

Finalmente, se calculan las proyecciones de las caras en el sensor de la imagen:

- $\tan(\beta) = \frac{x}{f} \Rightarrow x = f \cdot \tan(\beta)$
- $\tan(\beta + \gamma) = \frac{x + C_L}{f} \Rightarrow C_L = f \cdot \tan(\beta + \gamma) - x$
- $\tan(\beta + \gamma + \delta) = \frac{x + C_L + C_S}{f} \Rightarrow C_S = f \cdot \tan(\beta + \gamma + \delta) - x - C_L$

De los resultados algebraicos obtenidos se puede tener una expresión cerrada de  $L_1$ ,  $L_2$ , y el cociente  $K = \frac{C_L}{C_S}$ , para cada eje de la cámara  $x, y$ . Esto permite calcular el terreno total capturado por cada imagen, y hasta qué pixel se puede asumir una perpendicularidad total de los cuerpos captados para cualquier imagen agrícola que posean información geoespacial (GPS) y de rotación (IMU) de la cámara; las características intrínsecas de la misma; y una descripción cuantitativa aproximada de las plantas en cuestión.

## 3.2. Preparación de las fotografías

Las fotografías deben pasar por un proceso de recorte vertical y horizontal para los sectores que no cumplan con el criterio de perpendicularidad. Para ello se plantea que si  $K_{max}$  es la razón máxima que se aceptará entre la cara lateral ( $C_L$ ) y la cara superior ( $C_S$ ) captadas, entonces todo árbol de altura  $H_A$  y ancho  $W_A$  que se encuentre a una distancia  $R \in [-L_1, L_2]$

y que no cumplan esta condición, serán eliminados de la fotografía. Para lograr esto se busca el intervalo de  $R$  que cumpla con el criterio de perpendicularidad, cálculo que se realiza en un proceso numérico iterativo, ya que no es posible despejar esta variable del conjunto de ecuaciones obtenidos como para encontrar el valor teórico, calculando el cociente  $K$  para diferentes valores  $R_i$  discretizando el intervalo en  $n$  espacios equidistantes siendo el  $i$ -ésimo valor calculado como  $R_i = \frac{(i-1) \cdot (L_1 + L_2)}{n-1} + L_1$ .

Luego de obtener el intervalo correcto se procede a eliminar las porciones de la fotografía que no cumplan con el criterio de perpendicularidad. Para ello se utiliza un sistema de máscaras, donde todas las regiones de la fotografía que no serán utilizadas pasarán a tener el valor 0 en esta matriz, y 1 las que si. Esto se hace, ya que en los pasos posteriores se necesitan las fotografías completas para realizar ciertos cálculos. Continuando, sea  $P_x$  y  $P_y$  la cantidad total de pixeles en cada eje, y  $R_{x1}$ ,  $R_{x2}$ ,  $R_{y1}$ , y  $R_{y2}$  los límites inferiores y superiores de los intervalos calculados anteriormente para cada eje. Con esto en consideración, para obtener los pixeles que deben ser cambiados en la máscara según cada eje se debe realizar un mapeo entre la longitud del terreno y sus pixeles correspondientes:  $p(r) = \lfloor \frac{P_i}{L_1 + L_2} \cdot r + \frac{P_i \cdot L_1}{L_1 + L_2} \rfloor$ , donde  $p$  corresponde al pixel asociado a la longitud  $r$ , por lo que se puede saber de forma exacta cuáles son los pixeles de la imagen que deben mantenerse y los que no, algo posible utilizando los límites inferiores y superiores mencionados.

Con la imagen ya ajustada, se procede a encontrar cuáles son las imágenes sobrepuestas que comparten regiones del cultivo. Al contar con la información geoespacial y las dimensiones de la fotografía recortada, se puede conocer cuáles son las imágenes sobrepuestas utilizando las coordenadas de sus vértices. Una imagen se encuentra solapada con otra, siempre y cuando, a lo menos uno de los vértices de una fotografía se encuentre dentro del rectángulo formado por la otra 3.2. Un algoritmo sencillo para determinar si un vértice (un punto) se encuentra dentro de un rectángulo es trazando un rayo a partir de éste. Si el rayo interseca con un solo lado, el punto se encuentra dentro del rectángulo, y si interseca en dos o en ninguno, el punto se encuentra fuera 3.3.

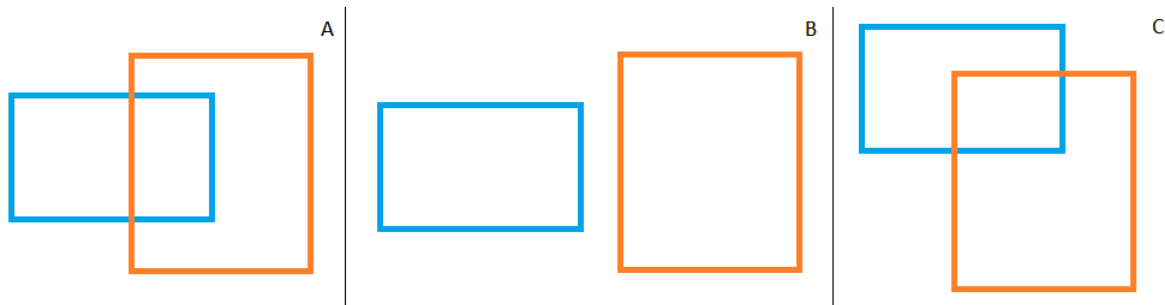


Figura 3.2: Representación de tres casos que se pueden dar entre dos fotografías al momento de verificar el solapamiento entre ellas. Caso A: Una de las fotografías posee dos de sus vértices dentro de otra. Caso B: Las fotografías no se encuentran solapadas. Caso C: Una de las fotografías posee un vértice dentro de otra. Cualquier otro caso de solapamiento no es considerado, debido a la naturaleza de la toma de fotografías con drones.

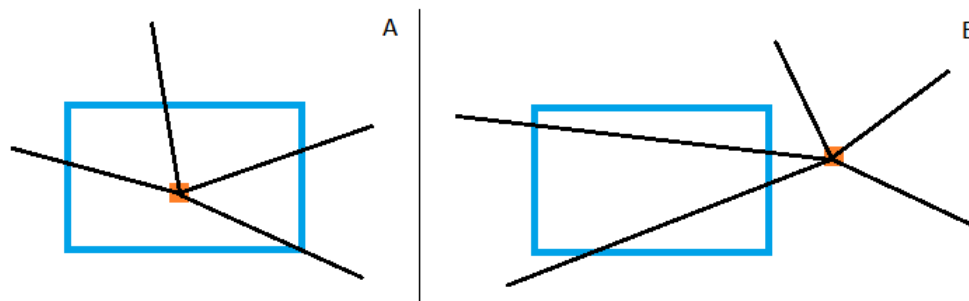


Figura 3.3: Muestra de los dos casos que se pueden dar cuando se quiere determinar si algún vértice de un rectángulo se encuentra dentro de otro. Caso A: Si un vértice de un rectángulo se encuentra dentro de otro, al trazar cualquier rayo desde el punto intersecará en una sola arista del rectángulo. Caso B: Si Un vértice se encuentra fuera del rectángulo, al trazar cualquier rayo desde el punto intersecará con ninguno o con dos aristas del rectángulo.

Utilizando el procedimiento descrito en esta sección, se cuenta con un conjunto de fotografías donde todos sus objetos cumplen con el principio de perpendicularidad, y que además se conoce cuáles son las imágenes que comparten información entre sí. El siguiente paso, descrito en la próxima sección, consiste en encontrar con mayor detalle la información compartida entre los pares de fotografías ya encontrados. Esto tiene el propósito de poder sobreponer las imágenes de manera más precisa unas con otras, como primer ajuste a la información geoespacial que éstas traen por defecto.

### 3.3. Búsqueda de puntos comunes entre fotografías

Los algoritmos realizados en la sección anterior, además de permitir que las imágenes estén bien adecuadas para la generación del ortomosaico, están pensados para reducir la cantidad de procesamiento que se tendría que hacer en los pasos posteriores. El paso siguiente consiste en encontrar la relación que poseen las imágenes entre si de manera más detallada, interesando cuáles son las que se encuentran solapadas. Para ello se detectan los puntos característicos de cada imagen; se emparejan estos entre pares de imágenes; y finalmente se realiza una verificación para eliminar aquellos emparejamientos erróneos. Todo esto sería muy costoso computacionalmente si no se tuviese inicialmente un emparejamiento de imágenes adyacentes obtenidos por la información geoespacial de la cámara, ya que el emparejamiento de puntos claves tendría que realizarse entre todas las imágenes, algo no necesario si ya se tiene una primera aproximación [42].

#### 3.3.1. Detección de puntos claves

El proceso de detección de puntos clave corresponde a una identificación automática de múltiples características primitivas o atributos del conjunto de imágenes tomadas sin la necesidad de requerir especificaciones de algún objetivo inicial. Algunas características son naturales, en el sentido que dichas características están definidas por la apariencia visual de

la imagen, mientras que otras son características artificiales resultantes de alguna manipulación específica realizada en la imagen (Figura 3.4). En el área del procesamiento de imágenes y visión computacional es necesario representar la fotografía por las características extraídas de ella, ya que la imagen en bruto permite que el ojo humano pueda extraer información sin embargo, no es el caso de los algoritmos computacionales [43] [44].

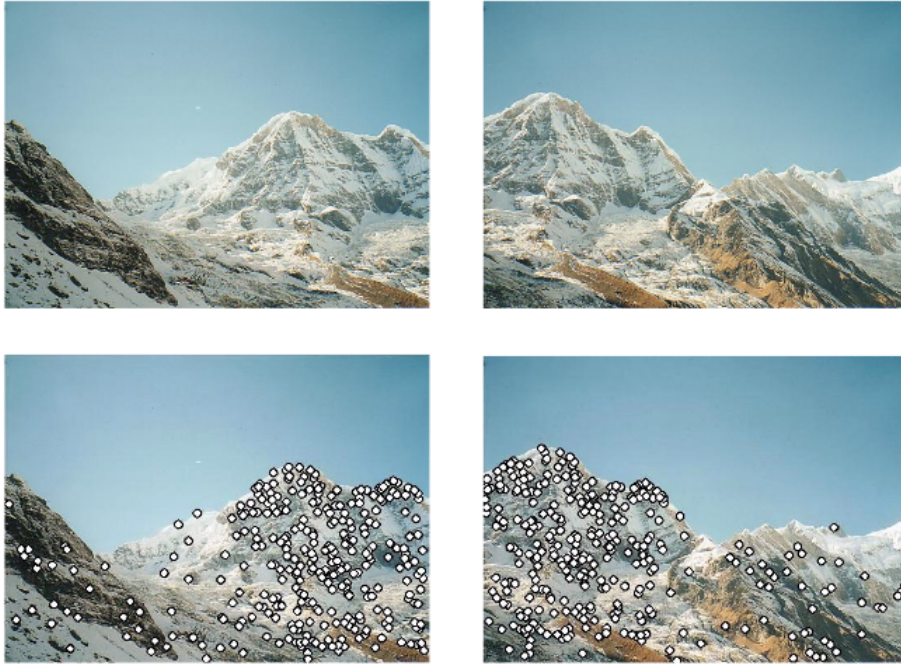


Figura 3.4: Par de imágenes en que se quiere verificar si están solapadas. Con el algoritmo SIFT se encuentra un conjunto de puntos repartido a lo largo de cada fotografía [14].

Existen generalmente dos métodos para representar imágenes: características globales, donde la imagen es representada por un vector multidimensional de características que describe la información de la imagen completa; y características locales, que representan la imagen basado en algunas regiones sobresalientes (puntos clave) mientras que permanecen invariantes a los cambios del punto de vista y de la iluminación. El detector y extractor de características debe poseer ciertas propiedades para que sea de utilidad, como robustez; repetibilidad; exactitud; generalidad; eficiencia; y cantidad. Una vez que un conjunto de puntos de interés ha sido detectado en una imagen en cierta posición, escala y orientación, su contenido o estructura necesita ser codificada en un descriptor adecuado para realizar emparejamientos discriminativos e insensible a deformaciones locales de la imagen [15].

La lista de métodos existentes para detectar puntos claves en imágenes es bastante extensa, sin embargo, el algoritmo SIFT es aquel que presenta mejor rendimiento según el estado del arte [42] [45] [46] [47] [48] en la búsqueda de puntos comunes en imágenes aéreas sobrepuestas. Los resultados positivos que este tiene se asocian a su invarianza tanto de rotación como de escala. Además de las ubicaciones de los puntos clave, SIFT proporciona un descriptor local para cada punto clave, haciendo que una imagen contenga varios miles de puntos clave SIFT como se explicó en el marco teórico.

### 3.3.2. Emparejamiento de puntos claves

Una vez que se tienen identificados los puntos característicos y representados por descriptores, para cada par de imágenes que comparten información, según ya fue identificado en la sección anterior, se buscan correspondencias entre si, proceso llamado emparejamiento de puntos claves. El objetivo de esta fase es emparejar los puntos encontrados de una imagen con los de las demás, utilizando como comparador los vectores de características (Figura 3.5). Esto permite inferir qué imágenes están solapadas e identificar así qué puntos corresponden a la misma información de la escena fotografiada. Para este fin se define una medida de distancia entre los descriptores de los puntos de interés y generalmente, el desempeño del método depende tanto de las propiedades del punto como de la elección del descriptor de la imagen [49] [15].

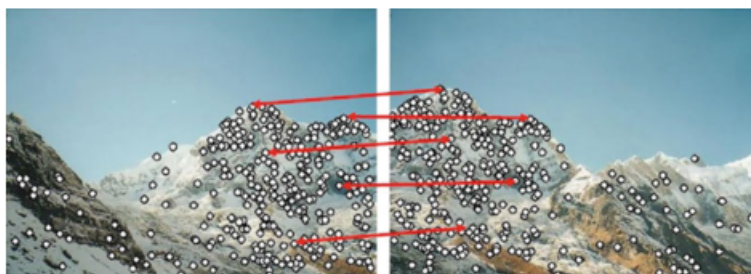


Figura 3.5: Emparejamiento de puntos claves entre dos imágenes basado en descriptores locales [15].

Para abordar este problema se define  $F(I)$  como el conjunto de características encontradas en la imagen  $I$ . Para cada par de imágenes  $I$  y  $J$ , el sistema considera cada descriptor  $f \in F(I)$  y encuentra su vecino más cercano (en el espacio de los descriptores)  $f_{nn} \in F(J)$  como:

$$f_{nn} = \operatorname{argmin}_{f_j \in F(J)} \|f - f_j\|_2$$

Para encontrar el argumento que minimiza la expresión se calcula la distancia euclidiana entre cada par posible, obteniendo así el emparejamiento entre ambos descriptores. Luego de emparejar las características de  $I$  y  $J$ , cada característica  $f \in F(I)$  será emparejada con a lo más una característica en  $F(J)$ , y algunos de estos emparejamientos pueden estar equivocados. Si luego de este proceso, un par de imágenes posee menos que un mínimo de emparejamientos, las imágenes se consideran que no calzan, y todos los emparejamientos son removidos de estas. De todas formas, es necesario un procedimiento más detallado para encontrar falsos positivos, el que será descrito a continuación [50].

### 3.3.3. Eliminación de emparejamientos erróneos

Se tienen los emparejamientos, de los cuales muchos son falsos positivos debido a la naturaleza del procedimiento realizado. Para reducir lo mayor posible los emparejamientos erróneos, se aprovecha el concepto de homografía entre pares de imágenes como se describió en el marco

teórico [34] [35]. Para ello se busca una homografía de primera aproximación entre cada par de imágenes que tengan emparejamientos de puntos. Esto, al ser un mapeo entre un plano y otro, aquellos pares de puntos que tengan un desvío o error muy grande se pueden considerar emparejamientos erróneos. Para realizar esta estimación se utiliza el algoritmo RANSAC.

Para implementar el algoritmo RANSAC en este problema se utilizan los emparejamientos realizados en el paso anterior del desarrollo. Se seleccionan cuatro correspondencias de forma aleatoria, y se resuelve el sistema de ecuaciones mencionado anteriormente, con lo que se puede obtener la matriz  $H$  sin problema [51]. Después se calculan la cantidad de inliers como se explicó anteriormente, proceso que se itera un número determina de veces, y se termina seleccionando la matriz de homografía que logra los mejores resultados. El proceso termina eliminando una gran parte de los outliers (emparejamientos erróneos) de todos los pares de imágenes que comparten información, algo que será aprovechado en el siguiente paso (Figura 3.6) [47].

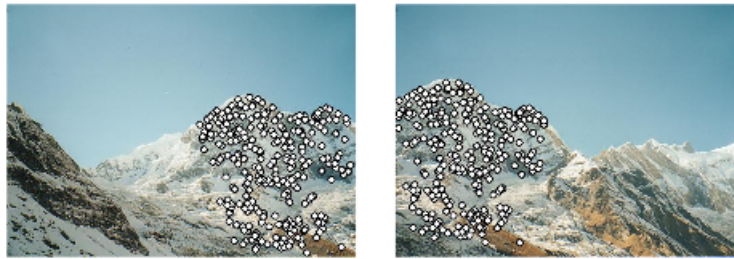


Figura 3.6: Emparejamientos mantenidos luego de realizar el algoritmo RANSAC sobre los emparejamientos encontrados. Con esto se borran la mayor parte de los falsos positivos emparejados [16].

Por otra parte, con esto se consigue una primera aproximación de las matrices de homografía entre los pares de imágenes que poseen regiones comunes (solapadas), lo que permite que una puede ser completamente mapeada en la otra (Figura 3.7). El problema que tiene esto, es que cada una de las homografías encontradas están relacionadas a una imagen distinta, es decir, a un plano diferente, por lo que es necesario utilizar estos mismos emparejamientos, pero encontrar las transformaciones homográficas con respecto a un mismo plano, cosa de que puedan ser mapeadas en éste y que se genere el ortomosaico. Si bien se puede pasar de plano a plano concatenando las homografías, este mecanismo termina propagando también los errores de las mismas transformaciones, ya que éstas son solo aproximaciones.



Figura 3.7: Dos imágenes que realizando una transformación homográfica adecuada quedan superpuestas [16].

### 3.4. Mejoramiento de la matriz de homografía

El objetivo de esta sección es estimar las matrices de homografía considerando todos los emparejamientos por imágenes y referenciando a un mismo plano. Para ello se considera que se encontraron  $N$  emparejamientos en las  $M$  imágenes del conjunto. Luego, el emparejamiento  $i$ -ésimo posee dos imágenes asociadas; dos puntos emparejados; y dos homografías al plano de referencia:  $I_a(i), I_b(i), p_a(i), p_b(i)$ , y  $H_a(i), H_b(i)$  respectivamente. El plano de homografía queda definido por la imagen de referencia  $I_{a,b}(ref)$  aquella que mejor cumpla el criterio de perpendicularidad, y cuya matriz de homografía es igual a la identidad  $H_{a,b}(ref) = I$  para que no sea modificada, con el objetivo de que el ortomosaico se vaya creando sobre el mejor plano posible.

Con esto en consideración los parámetros de cada matriz de homografía  $H_i \in \{H_1, \dots, H_M\}$  se escriben en un vector  $H_i = [H_{i11}, H_{i12}, H_{i13}, H_{i21}, H_{i22}, H_{i23}, H_{i31}, H_{i32}, H_{i33}]^T$ , vectores contenidos en la matriz  $H = [H_1, \dots, H_M]$  el siguiente sistema de optimización:

$$H^* = \underset{H_1, \dots, H_M \in H}{\operatorname{argmin}} \sum_{i=1}^N \frac{1}{2} \|e_i\|^2 = \frac{1}{2} e_i^T e_i \quad (3.3)$$

donde  $e_i$  corresponde a la distancia euclidiana del mapeo del emparejamiento  $i$ -ésimo cuyas coordenadas homogéneas son  $P_a$  y  $P_b$  en sus imágenes correspondientes  $I_a(i), I_b(i)$ , realizada con las matrices de homografías  $H_a(i)$  y  $H_b(i)$  sobre el plano referencial, ya que estos debiesen quedar en el mismo pixel (Figura 3.8). La diferencia queda expresada matemáticamente como:

$$e_i = C(H_a(i) \cdot P_a(i)) - C(H_b(i) \cdot P_b(i)) \quad (3.4)$$

La función  $C(p)$ , donde  $p = (X, Y, W)$ , realiza la conversión de la coordenada homogénea a la coordenada estándar del sistema de referencia del plano escogido. Para esto la función realiza el siguiente cálculo:

$$C(p) = \left( \frac{X}{W}, \frac{Y}{W} \right)$$



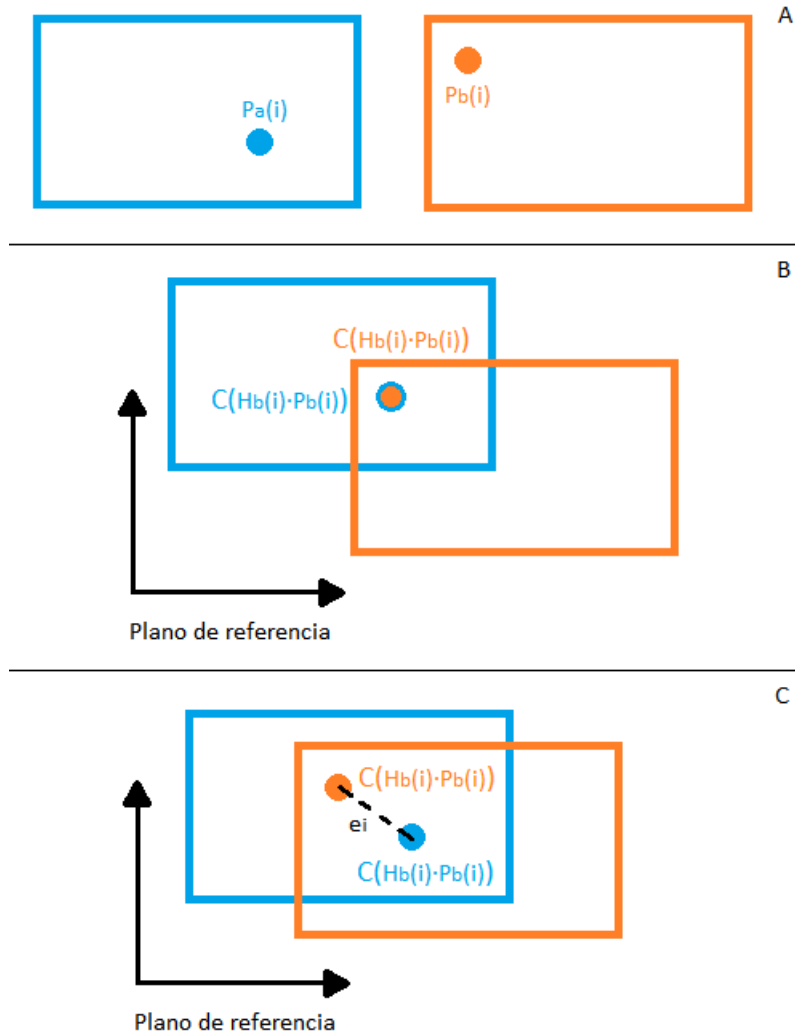


Figura 3.8: Representación gráfica de la diferencia entre la proyección del emparejamiento  $i$ -ésimo sobre el plano de referencia. Caso A: Se muestran ambas fotografías involucradas en el emparejamiento  $i$ -ésimo con sus puntos correspondientes. Caso B: Se muestra el caso en que la matriz de homografía hacia el plano de referencia fuese perfecta, calzando en el mismo pixel ambos puntos, siendo el error de proyección  $\|e_i\|^2 = 0$ . Caso C: Se muestra el caso en que las proyecciones no logran encontrarse en el mismo pixel, obteniéndose un error  $\|e_i\|^2 > 0$ .

Para resolver este sistema de optimización hay que encontrar los  $8(M - 1)$  parámetros que minimizan la expresión. Al ser un problema no lineal de mínimos cuadrados [52], se resuelve sin problemas con el algoritmo Levenberg-Marquardt (LM), un método descendiente donde en cada paso se minimiza la función objetivo, siendo capaz de encontrar el mínimo incluso si la solución inicial se encuentra lejana de la respuesta, como indica el trabajo [53]. Este algoritmo exige que se proponga una solución inicial, para lo que se utilizan las homografías encontradas en la sección anterior. Cada una de ellas es multiplicada de manera recursiva hasta llegar al plano de la imagen de referencia, es decir,  $H_F = H_{ref-a} \cdot H_{a-b} \dots$ , si la homografía de la imagen  $b$  hace referencia a la  $a$ , y si la imagen  $a$  hace referencia a la imagen de referencia, proceso que debe llevarse a cabo las veces que sea necesario. Esto se puede

realizar, ya que cada punto puede ser llevado a planos concatenados de la siguiente forma:

$$P'_a = H_{a-b} \cdot P_b \quad (3.5)$$

$$P'_{ref} = H_{ref-a} \cdot P_a \quad (3.6)$$

$$\Rightarrow P'_{ref} = H_{ref-a} \cdots H_{a-b} \cdot P_b \quad (3.7)$$

$$\Rightarrow P'_{ref} = H_{ref-b} \cdot P_b \quad (3.8)$$

Realizando el procedimiento descrito, se obtiene la estimación de la matriz  $H$  que posee todas las homografías de las imágenes, permitiendo mapear todas éstas a un solo plano, creando imágenes sobrepuestas y teniendo la primera versión del ortomosaico.

## 3.5. Mejoras al aspecto del ortomosaico

### 3.5.1. Ajuste de intensidades de las fotografías con ganancias variables

Las imágenes con regiones sobrepuestas, como se logró encontrar en la sección anterior, poseen en casi todos los casos diferencias de intensidades a pesar de referirse al mismo punto, lo que hace parecer que se vean los bordes de las fotografías y que el ortomosaico no sea homogéneo. Si se quiere arreglar esto y que los bordes de las fotografías no sean perceptibles, las intensidades de pixeles superpuestos deben ser lo más similares posibles. Para ello, se utiliza el algoritmo descrito en el marco teórico que es capaz de adecuar las intensidades de las imágenes, según una ganancia variable por fotografía (Figura 3.9) [14].



Mosaico sin compensación de ganancia



Mosaico con compensación de ganancia

Figura 3.9: Diferencia de los resultados de un mismo mosaico cuando se realiza un ajuste de intensidades de las fotografías con ganancias variables [14].

### 3.5.2. Mezcla de imágenes en múltiples bandas

Incluso luego de la compensación de intensidades mediante ganancias variables, algunos bordes de las imágenes siguen siendo visibles por una serie de efectos no deseados, como el vignetting; parallax effects; distorsiones radiales; entre otros. Es por todo esto que se debe implementar una buena estrategia para mezclar las fotografías. Para ello se usa la técnica llamada Mezcla de imágenes en múltiples bandas, enfoque descrito en el marco teórico [14].

El método descrito considera que la máscara de pesos debe realizarse de manera lineal, lo que daría como resultado una distribución circular, una aproximación que podría mejorarse. Para ello se propone generar una distribución no lineal que dé como resultado una distribución elipsoidal, algo que se aproxime mejor al efecto que éste fenómeno tiene sobre las fotografías (Figura 3.10)

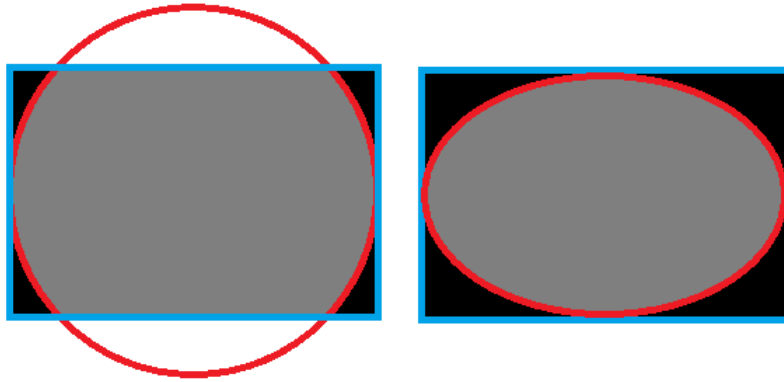


Figura 3.10: Comparación de la distribución obtenida entre una función lineal (imagen izquierda) y una no lineal (imagen derecha). La última corresponde a una función que genera una distribución elipsoidal proporcional a las correspondencias de la imagen.

Las elipses generadas en la máscara serán proporcionales a la más grande que se puede tener inscrita dentro de la imagen, es decir, su semieje mayor del largo de la imagen ( $W$ ), y semieje menor del ancho ( $H$ ), obteniendo la relación  $W = K \cdot H$ . Con esto en consideración, la ecuación de la elipse:

$$\frac{x^2}{w^2} + \frac{y^2}{h^2} = 1 \quad (3.9)$$

Se convierte en:

$$\frac{x^2}{(K \cdot h)^2} + \frac{y^2}{h^2} = 1 \quad (3.10)$$

Por lo tanto, según la coordenada del punto, considerando el centro de la imagen como origen, se puede obtener a la elipse que corresponde desarrollando la expresión anterior:

$$h = \sqrt{\frac{x^2 + K^2 y^2}{K^2}} \quad (3.11)$$

Luego, la asignación de pesos se realiza considerando las condiciones de borde, donde la elipse de semieje menor más pequeña ( $h_{min}$ ) le corresponde el valor de 1, y la más grande ( $h_{max}$ )

se le asocia el 0. Generando la siguiente ecuación de la recta:

$$M(h) = -\frac{1}{h_{max} - h_{min}} \cdot h + 1 \quad (3.12)$$

Reemplazando por las expresiones conocidas en la ecuación de la recta se obtiene la nueva matriz de pesos:

$$M(x, y) = -\sqrt{\frac{K^2}{W^2 + K^2 H^2}} \cdot \sqrt{\frac{x^2 + K^2 y^2}{K^2}} + 1 \quad (3.13)$$

Todo el resto del método se desarrolla como explica el marco teórico.

## 3.6. Generación de métricas para medir el desempeño del algoritmo

La calidad de cualquier ortomosaico generado puede ser evaluada por tres grandes parámetros: Si posee artefactos visuales a lo largo de la imagen resultante; la precisión geométrica del ortomosaico; y el costo computacional que tiene el algoritmo para obtener los resultados. En esta sección se explica en mayor detalle a qué se refiere cada aspecto, cómo se miden, y qué implicancias tienen [54].

### 3.6.1. Artefactos visuales

Un artefacto corresponde a una parte del contenido de una imagen que no tiene un equivalente en el objeto físico que se está visualizando, lo que da a entender la existencia de estructuras falsas. Es distinto del ruido, ya que éste corresponde a una contribución a la señal en cuestión causada por procesos aleatorios, de naturaleza estadística y que no contienen información sobre la señal [55].

La cantidad de artefactos presentes en un ortomosaico pueden aparecer en cualquiera de los pasos de la metodología propuesta para su construcción. Es importante que los artefactos sean nulos, sin embargo, en caso de que existan, que sean muy pequeños y en baja cantidad. Al tener artefactos en los resultados puede llevar a interpretaciones erróneas o a dificultar las mismas, motivo por el que al momento de generar el ortomosaico se analizarán en detalle. Este aspecto es evaluado mediante la apreciación visual por un usuario experto.

### 3.6.2. Precisión geométrica

Las fotografías utilizadas poseen la información GPS de cada una, permitiendo seleccionar manualmente una serie de píxeles, repartidos a lo largo del ortomosaico, como puntos de control. Para ello se seleccionan puntos característicos del ortomosaico y se obtiene su coordenada terrestre, logrando tener un arreglo de posiciones en las que debiesen estar los puntos, permitiendo comparar referencialmente los generados. Para ello basta con encontrar

aquellos puntos seleccionados en el ortomosaico y calcular la diferencia entre la coordenada resultante, y la medida por el GPS, permitiendo tener una métrica de error. Con este objetivo, se utiliza el RMSE (Root-Mean-Square Error); el error más pequeño encontrado; y el error más grande [54]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{N}} \quad (3.14)$$

$$e_{max} = \arg \max_i |(\hat{y}_i - y_i)| \quad (3.15)$$

$$e_{min} = \arg \min_i |(\hat{y}_i - y_i)| \quad (3.16)$$

donde  $(\hat{y}_i - y_i)$  corresponde al error existente entre la coordenada de control obtenida con el GPS, y la coordenada obtenida por el ortomosaico;  $e_{max}$  es el error más grande obtenido; y  $e_{min}$  el menor.

### 3.6.3. Costo computacional

Al momento de analizar el costo computacional, es importante probar para un número variable de imágenes y distintos conjuntos de fotografías el tiempo que toma obtener el resultado final. Con esto se puede ver la complejidad computacional y estimar el orden de magnitud del tiempo de ejecución, e incluso, si se toman los tiempos por etapas, se pueden ver qué partes de la metodología toman más tiempo que las demás, como para proponer optimizaciones de recursos y que el algoritmo se ejecute en menor tiempo [54].

Considerando que las áreas agrícolas son muy grandes, la cantidad de fotografías para un ortomosaico de estas características son muy elevadas. Además, los ortomosaicos al estar pensados para tomar mejores decisiones en los cultivos, estos deben estar listos lo antes posible. Y como última acotación, mientras mejor sea el orden de los algoritmos, la computadora necesaria para generar los resultados puede ser de menor costo, reduciendo la inversión que se deba realizar.

# Capítulo 4

## Resultados

La metodología expuesta fue implementada en el lenguaje Matlab en un flujo lineal, permitiendo analizar y ver los resultados obtenidos en cada una de las etapas. El computador utilizado para realizar las pruebas posee un procesador Intel(R) Core(TM) i7-2600 CPU 3.40GHz, 12 GB de memoria RAM instalada, sistema operativo Windows 8 de 64 bits, y un disco de estado sólido de 128 GB. Como se mencionó en el capítulo anterior, los resultados serán analizados desde tres perspectivas según vaya al caso, análisis y presentación de artefactos visuales, precisión geométrica en el ortomosaico construido, y el costo computacional de la etapa en cuestión. El desempeño es medido utilizando dos bases de datos con fotografías aéreas de la empresa senseFly, dedicada a la venta de drones especializados para la toma de fotografías aéreas. En su sitio web la empresa ofrece varias bases de datos tomadas por sus vehículos aéreos no tripulados. De todo este conjunto, se utilizan dos, una llamada "High precision mapping of a technology park", tomada sobre el parque de tecnología más largo de Suiza; y la otra llamada "Assising crop with RGB imagery (eBee SQ)", fotografías de un cultivo ubicado en el mismo país [56].

### 4.1. Criterio de perpendicularidad

Utilizando los metadatos de las bases de datos mencionadas se logra acotar las dimensiones de las fotografías para que se cumpla el criterio de perpendicularidad descrito en el capítulo 3. Para visualizar de mejor forma los efectos del recorte de las imágenes, se utilizará la primera base de datos, ya que posee edificios, los tienen un aspecto simple y regular. La altura de vuelo promedio de cada fotografía es de 100 metros. Asumiendo que los edificios son de tres y dos pisos, se puede suponer un alto y ancho aproximado de 8 y 15 metros, respectivamente. Si se considera una constante de perpendicularidad del 0.05, se obtienen transformaciones como muestra la Figura 4.1, donde a la izquierda se tiene la imagen original y a la derecha la ajustada. De aquí se puede ver que el edificio de más a la izquierda alcanza a ser eliminado de la imagen, en vista de que, según el criterio de perpendicularidad, una gran parte de su cara lateral es fotografiada. No así el caso de la Figura 4.2, donde se puede ver que la región eliminada es mucho menor.



Figura 4.1: Ajuste que recibe una fotografía urbana (imagen de la izquierda) al momento de redimensionarla para que cada uno de sus píxeles cumpla con el criterio de perpendicularidad (imagen de la derecha).



Figura 4.2: Ajuste que recibe una fotografía agrícola (imagen de la izquierda) al momento de redimensionarla para que cada uno de sus píxeles cumpla con el criterio de perpendicularidad (imagen de la derecha).

En este caso no amerita medir los costos computacionales que trae ejecutar el algoritmo, ya que además de ser muy bajos, es un procedimiento de orden lineal. Por otra parte, en las imágenes presentadas se puede ver que los recortes realizados para un criterio de perpendicularidad de 0.05 en el contexto de la agricultura, si la cámara no posee una rotación considerable, se conserva casi la gran mayoría de la información, ocurriendo un recorte de la imagen pequeño en comparación al caso del sector urbano.

## 4.2. Información del GPS e IMU en la búsqueda de puntos comunes

Las fotografías de las bases de datos descritas poseen información tanto GPS como IMU, requisito exigido por la metodología planteada. Luego de que éstas pasan por la eliminación de píxeles recién descrita, las fotografías son disminuídas en escala, ya que sus dimensiones son de  $5472 \times 3648$  píxeles y  $4608 \times 3456$  píxeles respectivamente de manera original, algo que aumenta considerablemente el costo de computación en los pasos siguientes si se trabaja con aquellas dimensiones. Es necesario acotar que mientras menor sea la disminución de escala efectuada, mayor es el detalle que se tiene de la fotografía, motivo por el que se debe tomar una decisión sobre cuál es el nivel de detalle que se desea tener en el ortomosaico final.

Durante el proceso de emparejamiento pueden ocurrir tres casos:

1. El primero corresponde al caso ejemplar, es decir, dos fotografías solapadas que compartan información sobre una misma escena y que son bien distinguibles una de otra. Generalmente en este escenario la cantidad de falsos positivos de pareamientos de puntos clave encontrados es relativamente bajo y no se tiene mayor problema para eliminarlos (Figura 4.3).

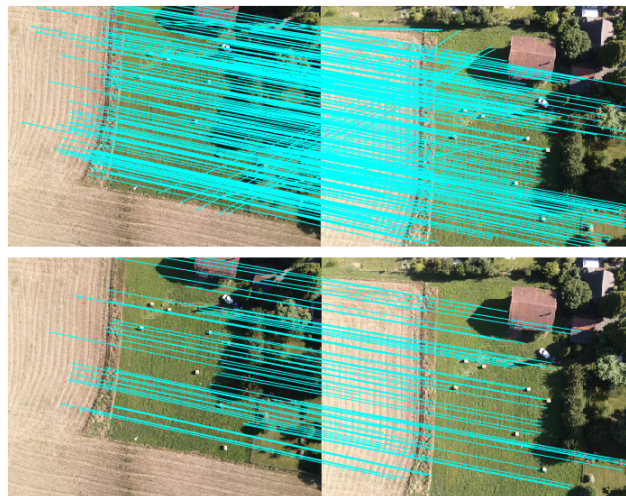


Figura 4.3: Emparejamiento realizado entre dos imágenes solapadas (fotografía de arriba), donde los emparejamientos mal realizados quedan completamente eliminados con el algoritmo RANSAC como se ve en la fotografía de abajo.

2. La segunda posibilidad es que las fotografías no estén solapadas, pero las imágenes son tan distintas una de otra, que los emparejamientos, en caso de encontrarse alguno, corresponden todos a falsos positivos, lo que hace que sea relativamente simple desechar que las imágenes se encuentran relacionadas (Figura 4.4).



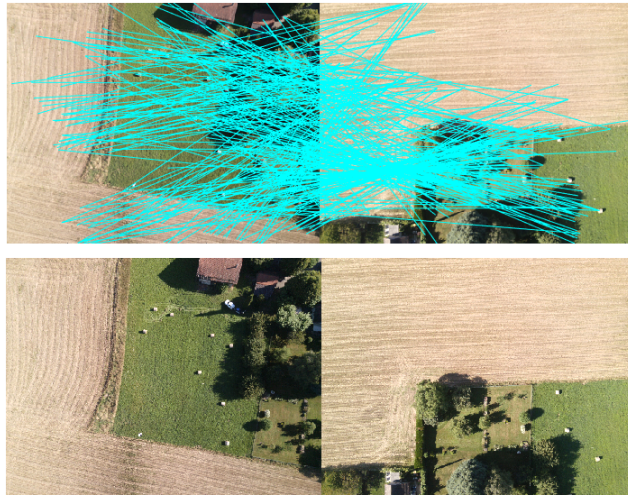


Figura 4.4: Emparejamiento realizado entre dos imágenes que no están solapadas (Fotografía de arriba). Los calces realizados son todos falsos positivos, los que con el algoritmo RANSAC quedan completamente eliminados, concluyendo que estas fotografías no comparten información de una escena (fotografía de abajo).

3. El último caso corresponde a fotografías muy similares, por ejemplo, una zona de cultivo con casi ningún punto de referencia como para caracterizar el espacio en que se ubica, lo que trae serios problemas al momento de realizar un emparejamiento (Figura 4.5). La dificultad que esto conlleva es que se podrían emparejar imágenes que no se encuentran realmente solapadas, debido a la similitud de características encontradas, o podrían desecharse emparejamientos que son reales.

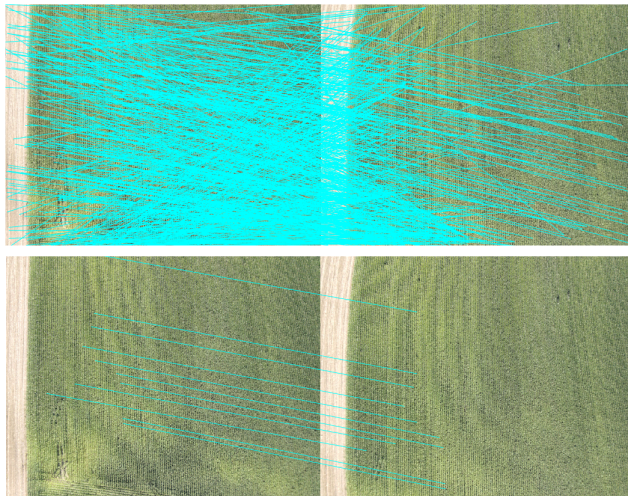


Figura 4.5: Emparejamiento realizado para dos fotografías muy similares que se encuentran solapadas. En la fotografía de arriba se puede ver la gran cantidad de emparejamientos, sin una dirección clara a simple vista, sin embargo, tras el algoritmo RANSAC, se logran encontrar un par de calces que explican el solapamiento entre ambas imágenes (fotografía de abajo).

Es en el último punto donde la información del GPS juega un papel fundamental, ya que al

conocer cuáles son las fotografías cercanas que están solapadas esta posibilidad prácticamente es eliminada.

Fue realizado un análisis sobre los tiempos de cómputo que trae este algoritmo considerando tres variables, el tiempo en que se desarrolla esta etapa, la escala a la que es transformada la imagen, y el número de imágenes a emparejar. Primero se mide el orden del algoritmo según distintas escalas para 50 conjuntos aleatorios de cuatro imágenes, donde mientras más grande sean sus dimensiones, es decir, menor la escala aplicada, mayor es la cantidad de puntos claves que se obtendrán para cada una, lo que hace que aumente el número de emparejamientos a realizar. Además, vale acotar que la escala produce un aumento lineal en cada dimensión espacial de la imagen, pero uno cuadrático en lo que respecta a la cantidad de píxeles, lo que explicaría los tiempos de cómputo promedios del experimento (Figura 4.6).

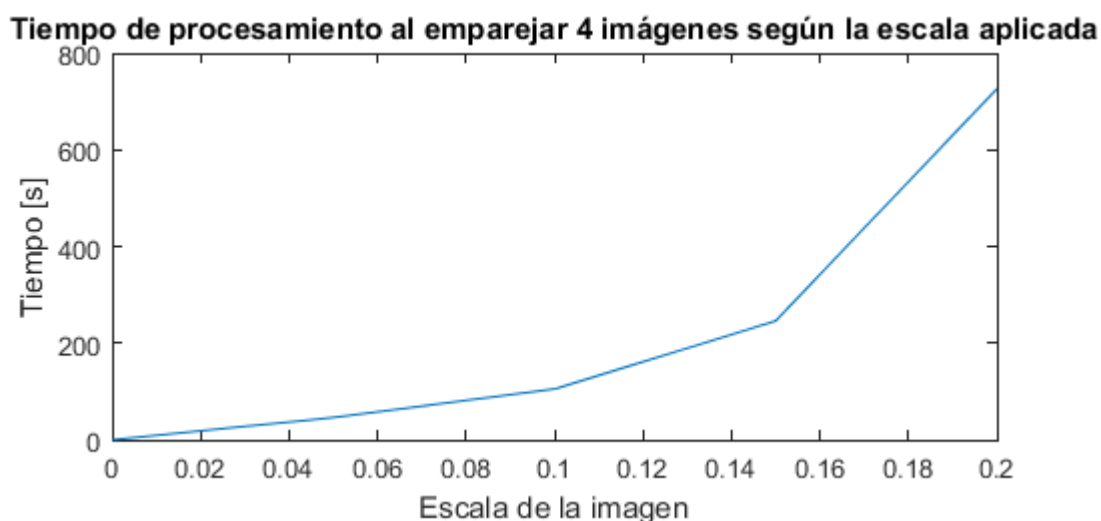


Figura 4.6: Gráfico que muestra el tiempo promedio de procesamiento que toma emparejar cuatro imágenes según distintas escalas aplicadas a las imágenes.

La otra prueba realizada consiste en medir los tiempos de cálculo que toma emparejar distintas cantidades de imágenes a una escala de 0.1 a cada una. Todas las fotografías utilizadas en esta verificación se encuentran solapadas con otras tres, por lo que cada imagen debe ser comparada solo con una cantidad fija gracias a que se utiliza la información GPS, teniendo un tiempo de cómputo lineal en lugar de cuadrático, ya que éstas no deben ser comparadas con el conjunto completo. Por otra parte, existen pequeñas fluctuaciones en los tiempos, ya que la cantidad de puntos claves varía en cada una, habiendo casos en que se deben comparar muchos menos descriptores a diferencia de otros. Al igual que en el experimento anterior, se tomaron de manera aleatoria 50 conjuntos de imágenes, calculando los tiempos de cómputo promedio (Figura 4.7)

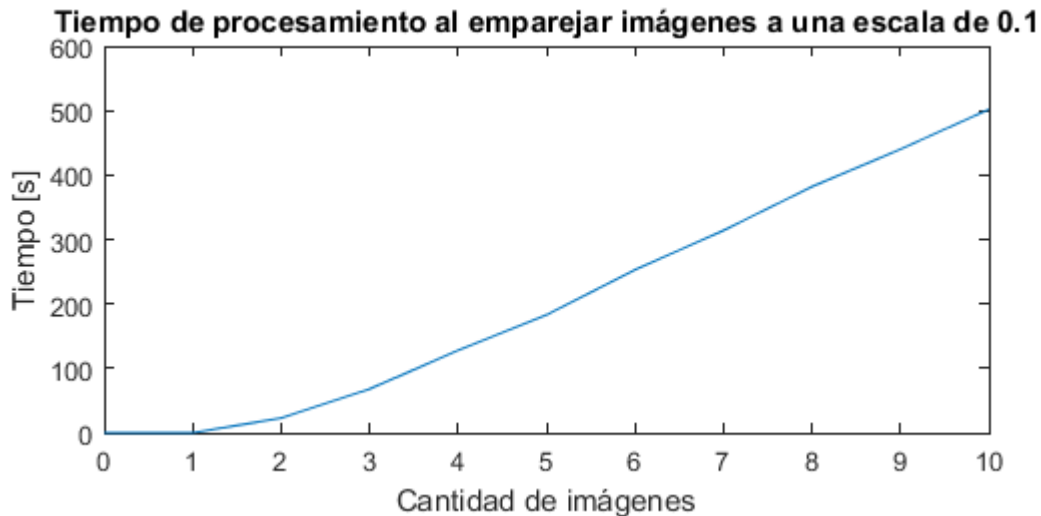


Figura 4.7: Gráfico que muestra el tiempo promedio de procesamiento que toma emparejar distintas cantidades de imágenes para una escala fija de 0.1.

### 4.3. Estimación de las matrices de homografía con respecto al plano de referencia

La estimación de las matrices de homografía con respecto al plano de referencia, y por ende la primera versión del ortomosaico, corresponde en gran parte al reflejo de lo que se ha hecho anteriormente. Como se vio en capítulos anteriores, lo único que se utiliza para encontrar las matrices son los puntos claves emparejados entre imágenes, por lo que tanto la calidad, como la cantidad, afectan directamente el resultado obtenido. Si la cantidad de emparejamientos es muy baja, lo más probable que ocurra es que las estimaciones no sean del todo correctas, resultando en mosaicos, debido a la distorsión geométrica que presentarían, y no a un ortomosaico como es lo que se busca en este trabajo. Por ejemplo, en la Figura 4.8, se puede ver un mosaico que a medida que éste va subiendo, la vista comienza a distorsionarse, dando un efecto de ladeo en la imagen, a pesar de que está tomada de forma perpendicular (Figura 4.9). Vale acotar, que otra de las causas de los resultados de baja calidad, corresponde al algoritmo y los parámetros con que se resuelve el sistema de optimización, ya que mientras mayor sea la cantidad de emparejamientos, si bien es mayor la información que le entrega al sistema, también lo es en complejidad, ocurriendo casos en que el sistema tarda varias horas en poder encontrar una solución no necesariamente óptima. Este también es el principal motivo de que los resultados mostrados correspondan a 6 imágenes sobrepuestas como máximo, debido a que para un número mayor es necesario tomar quizá otra estrategia para resolver el sistema de optimización planteado, más que incrementar la capacidad de cómputo del equipo. Si bien el resultado mostrado es visible y apreciable, a medida que se le siguen agregando imágenes se distorsiona de tal forma, que no tiene utilidad alguna.



Figura 4.8: Ortomosaico defectuoso debido a la baja calidad y/o cantidad de emparejamientos encontrados en las etapas previas, junto a una falta de mayor procesamiento para llegar al óptimo de la estimación de homografías.



Figura 4.9: Imagen original que es deformada en el ortomosaico de la Figura 4.8.

Cuando se tiene una buena cantidad y calidad de emparejamientos, además de parámetros bien configurados para resolver el sistema de optimización, se consiguen ortomosaicos de buen aspecto en lo que respecta a relaciones y deformaciones geométricas según usuario experto, pudiendo ver que cada imagen se modifica adecuadamente para ofrecer la información en cuestión de la mejor forma posible. Para éste y los futuros experimentos se toman dos conjuntos de seis fotografías aleatorias, Escenario A y Escenario B respectivamente. En las Figuras 4.10 y 4.11 se muestran los dos ortomosaicos generados a partir de los dos escenarios. En las imágenes se pueden apreciar una serie de particularidades, la primera y más llamativa, corresponde a unas líneas negras que aparecen en varias partes, dando a entender que corresponden a los bordes de las imágenes. También se ve que unas fotografías poseen ligeras variaciones de colores o luminosidad a pesar de referirse al mismo punto, haciendo que se note aún más que existe una imagen sobrepuesta. Además, se puede observar que, en algunas regiones de borde, los píxeles no calzan de manera perfecta, mostrando información repetida o eliminando otras. Finalmente, se aprecia que la imagen más lejana (la de más arriba), en ambos casos se encuentra un poco más deformada que las demás hacia fuera de ésta misma, efecto causado porque el sistema de optimización no tiene con qué emparejar los puntos que no se encuentran solapados con otra imagen, por lo que ellos quedan sin ajustarse y el sistema solo se encarga de hacer calzar aquellos emparejados. Esto tiene como repercusión en que generalmente, las imágenes de borde de uno de estos ortomosaicos creados, deben ser eliminadas ya que no presentan la información como se espera de un ortomosaico.



Figura 4.10: Ortomosaico del Escenario A obtenido de buen aspecto según usuario experto.



Figura 4.11: Ortomosaico del Escenario B obtenido de buen aspecto según usuario experto.

De todos los resultados mostrados, se puede ver que éstos van creciendo en un sentido fijo, en lugar de ir expandiéndose en todas las direcciones. Esto se debe a una complicación que se tuvo al momento de implementar el algoritmo, relacionada con el mapeo del plano de una homografía a otra. En el capítulo anterior, se comentó que la solución inicial que es entregada al algoritmo para resolver el sistema de optimización utiliza las homografías de un plano a otro de manera concatenada. El problema ocurre cuando no se logra encontrar un camino directo hacia el plano de referencia fijado, ya que, para hacerlo en todos los casos, es necesario tener una estructura de grafo que sea capaz de recorrerse en ambas direcciones, cosa de llegar al plano de referencia cual sea la situación. Todo este problema, exige que en el sistema implementado sea necesario agregar las fotografías en el mismo orden en que fueron tomadas durante el vuelo, lo que hace que los ejemplos luzcan de esa forma.

#### 4.4. Cambios de aspecto del ortomosaico

Realizando en primer lugar el ajuste de las intensidades de cada imagen, se obtienen resultados sutiles a nivel general, balanceando los colores de las imágenes como se esperaba. Si bien realizar este paso no elimina los bordes negros característicos del paso anterior, la apariencia de cada ortomosaico por si solo mejora bastante según usuario experto (Figura 4.12 y 4.13). Otro punto a rescatar es que el tiempo de cómputo que toma en realizar este ajuste es muy bajo, al ser solo un sistema de ecuaciones lineales que crece en una dimensión por cada fotografía. Y por último, no cambia las posiciones espaciales de los píxeles que ya

fueron estimados en pasos anteriores, por lo que esta etapa no merece tener un tipo de arreglo o mejora futura de manera urgente.



Figura 4.12: Escenario A (izquierda) al que se le aplica un ajuste de intensidades en base a ganancias variables (derecha).

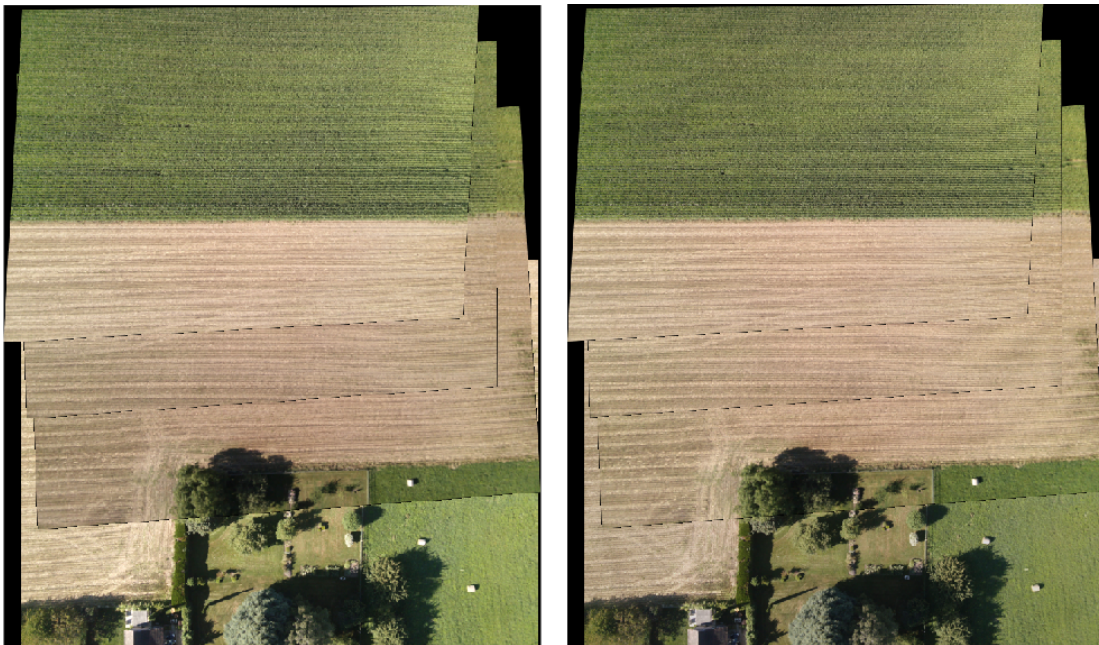


Figura 4.13: Escenario B (izquierda) al que se le aplica un ajuste de intensidades en base a ganancias variables (derecha).

Luego, con la mezcla de imágenes en múltiples bandas los resultados se observan de forma notoria (Figura 4.14 y 4.15), desapareciendo las líneas negras relacionadas a los bordes de las imágenes que se encuentran al centro del ortomosaico, uniendo detalles como la calle que en la versión anterior se encontraba cortada, árboles de los bordes que también estaban unidos de manera evidentemente mal, ahora tienen un aspecto homogéneo, lo mismo para el terreno de los cultivos, donde el detalle que se logra y el cambio son destacables. Sin embargo, lo que respecta a las secciones laterales del mismo, se pueden visualizar ciertas líneas generando artefactos y un claro error. Esto se debe a que en esas secciones del ortomosaico, no se tienen una imagen que represente mejor esa escena y debe tomar la información de las esquinas de las mismas, generando esos notorios problemas.



Figura 4.14: Escenario A al realizar una mezcla de imágenes en múltiples bandas.





Figura 4.15: Escenario B al realizar una mezcla de imágenes en múltiples bandas.

Los errores mencionados desaparecerían completamente si es que el ortomosaico creciera en todas direcciones, sin embargo, siempre en las últimas imágenes de cada borde del ortomosaico final, tendrá este tipo de errores, motivo por el que las fotografías deben ser tomadas un poco más lejanas del área de interés, con el fin de eliminar esas regiones del resultado final y obtener así una imagen homogénea de la zona agrícola que se desea estudiar.

Este recorte es realizado en los dos escenarios trabajados (Figura 4.16 y 4.17), y al ser formados por solo 6 imágenes, gran parte de la información se pierde, no obstante, se debe recordar que la sección a eliminar corresponde a los bordes de la misma. En una situación real, esto mismo debe ser aplicado para ortomosaicos formados por más de cien fotografías, por lo que eliminar los bordes hace que su tamaño siga siendo básicamente el mismo.



Figura 4.16: Escenario A al eliminar sus bordes tras realizar los ajustes de apariencia.



Figura 4.17: Escenario B al eliminar sus bordes tras realizar los ajustes de apariencia

Finalmente, se realiza un análisis de precisión en la distribución espacial que toma el ortomosaico al ir generándose. Para ello se tomaron manualmente 30 puntos de control, escogiendo aquellos sectores del ortomosaico que fuesen fáciles de reconocer, y que estén dispersos en la imagen. Las coordenadas de cada uno se obtienen con la misma relación utilizada en la primera fase, es decir, se ocupa la coordenada GPS central de la fotografía y se estima a cuál coordenada corresponde el pixel en determinada imagen. Con esto en consideración, se asume que las coordenadas de los cuerpos de la imagen de referencia están correctas y se realiza el mismo procedimiento, pero para obtener las coordenadas de los puntos de control en la versión final. A partir de esto es que se calcula el RMSE de las coordenadas obtenidas, el error más grande obtenido, y el más pequeño (Tabla 4.1). Los errores más grandes se obtuvieron en la imagen más lejana, ya que como se comentó anteriormente, éstas son las que poseen menos calces para ajustarse y por lo mismo son las que quedan sobreajustadas a los demás emparejamientos obteniendo mayores distorsiones. Sin embargo,

en las demás regiones del ortomosaico, los resultados son mucho mejores. Por otra parte, el hecho de que el RMSE sea ligeramente menor al promedio entre el error máximo y el error mínimo, se adjudica a que la cantidad de puntos de control por imágenes fue la misma, sabiendo que la última imagen es aquella que posee los errores mayores.

<b>Ortomosaico</b>	<b>RMSE [m]</b>	<b>Error mínimo [m]</b>	<b>Error máximo [m]</b>
Escenario A	3.8	1.2	10.2
Escenario B	1.9	0.4	4.3

Tabla 4.1: Errores espaciales obtenidos en ambos ortomosaico finales.

# Conclusión

La fotogrametría es la ciencia enfocada a la obtención de información sobre objetos físicos y el ambiente a través de fotografías, recurso de mucho valor para diferentes tipos de industrias, como reconocimiento y mapeo de sectores, minería, ingeniería y construcción, protección ambiental, y como se vio en este trabajo, la agricultura de precisión, área encargada de maximizar los recursos de terrenos agrícolas. Una de sus aplicaciones en la última área mencionada, corresponde a la posibilidad de generar ortomosaicos a partir de múltiples fotografías tomadas por vehículos aéreos no tripulados, para obtener una visión completa del terreno cultivado en diferentes bandas del espectro electromagnético, y así utilizar distintos índices y criterios para mejorar la toma de decisiones del campo.

Para efectos de este trabajo, y para poder cumplir con los objetivos planteados, se ha propuesto un algoritmo que genera ortomosaicos de manera automática para fotografías tomadas por drones de manera vertical que incluyan la información GPS e IMU, y sin la intervención de un usuario. Los beneficios de la metodología propuesta se basan principalmente en su simplicidad; modularidad; y facilidad de implementación y operación. Por otra parte, se puede apreciar que los ortomosaicos obtenidos sirven para varios de los propósitos de la agricultura de precisión, en vista de la homogeneidad y bajas distorsiones espaciales obtenidas.

Los resultados logrados a través del uso de la presente propuesta muestran que es escalable, ya que la mayoría de los pasos presenta cálculos de orden lineal, y para aquellos que no lo hacen, existen diferentes soluciones y herramientas que mejoran considerablemente su desempeño, por ejemplo, el árbol-kd para la comparación de descriptores. Además, al ser un sistema modular permite adecuar arquitecturas de trabajo especializadas para el trabajo en cuestión. Por otra parte, es una metodología que puede ser programada en variados lenguajes computacionales, lo que permitiría trabajar de manera más eficiente los recursos computacionales. También, es un algoritmo que puede ser paralelizado en varias etapas, como por ejemplo, en los recortes de las imágenes según su inclinación, en la obtención y comparación de descriptores y la transformación de las imágenes al plano de referencia. Por último, el sistema también es ampliable a imágenes de distintas bandas, no solo a las RGB, lo que permite que sea utilizable en una gran cantidad de aplicaciones. Todos los motivos mencionados indican que es una propuesta que puede ser concretamente aplicable al contexto real de la agricultura actual.

Según los análisis llevados a cabo, a esta metodología se le pueden realizar varias mejoras a corto plazo para perfeccionar tanto aspectos de precisión, como de visualización y órdenes de cómputo, listadas a continuación:

- Realizar una estructura de grafo para los solapamientos encontrados de las imágenes, junto a un algoritmo de recorrido como Dijkstra, para llegar al plano de referencia con la menor cantidad de concatenaciones de homografías.
- Generar un mejor algoritmo para la resolución del problema de optimización asociado a la estimación de las matrices de homografías de cada imagen hacia el plano de referencia, ya sea mejorando los parámetros del propuesto, o planteando uno completamente nuevo.
- Utilizar puntos de control sobre las fotografías para agregarlos al sistema de optimización recién mencionado como restricciones, lo que permitiría reducir los errores espaciales de las homografías.
- Mejorar el sistema de comparación de descriptores en la etapa de emparejamiento, tanto en órdenes de magnitud, como de umbrales variables para determinar si es correcto o no.

# Bibliografía

- [1] “National Cancer Institute, ‘Campos electromagnéticos y cáncer’, 2016. [En línea]. Disponible: [www.cancer.gov/espanol/cancer/causas-prevencion/riesgo/radiacion/hoja-informativa-campos-electromagneticos](http://www.cancer.gov/espanol/cancer/causas-prevencion/riesgo/radiacion/hoja-informativa-campos-electromagneticos). [Accedido: 11-Mar-2018].”
- [2] “Unmanned Tech Shop, ‘3 Axis Brushless Gimbal’, 2015. [En línea]. Disponible: [www.unmannedtechshop.co.uk/3-axis-brushless-gimbal-sony-nex-size-camera/](http://www.unmannedtechshop.co.uk/3-axis-brushless-gimbal-sony-nex-size-camera/). [Accedido: 11-Mar-2018].”
- [3] J. S. Aber, I. Marzoff, and J. B. Ries, *Small-format Aerial Photography: Principles, Techniques and Geoscience Applications*. Amsterdam: Elsevier, 2010.
- [4] “John A. Dutton e-Education Institute, ‘Clasificación de los Sistemas Aéreos no tripulados’, [En línea]. Disponible: <https://www.e-education.psu.edu/geog892/node/5> [Accedido: 11-Mar-2018].”
- [5] “Universität zu Köln, ‘BASP’, 2015. [En línea]. Disponible: [http://www.uni-koeln.de/~al001/radcor\\_files/rad027.png](http://www.uni-koeln.de/~al001/radcor_files/rad027.png). [Accedido: 07-Abr-2018].”
- [6] “Spojená škola, Juraja Henišča, ‘Krpáno’, 2016. [En línea]. Disponible: <http://ssjh.sk/krpano/examples/xml-usage/compass/vignetting.jpg>. [Accedido: 07-Abr-2018].”
- [7] “Brian Koberlein, ‘Aberration’, 2014. [En línea]. Disponible: <https://briankoberlein.com/2014/02/18/aberration/>. [Accedido: 07-Abr-2018].”
- [8] “BOOFCV, ‘Example Wavelet Noise Removal’, 2016. [En línea]. Disponible: [https://boofcv.org/images/1/18/Example\\_lena\\_denoise\\_noisy.jpg](https://boofcv.org/images/1/18/Example_lena_denoise_noisy.jpg). [Accedido 07-Abr-2018].”
- [9] “Physics Open Lab, ‘NDVI Index’, 2017. [En línea]. Disponible: <http://physicsopenlab.org/2017/01/30/ndvi-index/>. [Accedido: 11-Mar-2018].”
- [10] M. D. Pritt, “Fast orthorectified mosaics of thousands of aerial photographs from small UAVs,” *Proceedings - Applied Imagery Pattern Recognition Workshop*, vol. 2015-Febru, no. February, 2015.
- [11] “Learn OpenCV, ‘Image Alignment (ECC) in OpenCV C++/Python’, 2015. [En línea]. Disponible: <https://i.stack.imgur.com/IFC1t.jpg>.”

- [12] “Msm, ‘Data points of line with many outliers’. [En línea]. Disponible: [https://upload.wikimedia.org/wikipedia/commons/b/b9/Line\\_with\\_outliers.svg](https://upload.wikimedia.org/wikipedia/commons/b/b9/Line_with_outliers.svg). [Accedido: 11-Mar-2018],”
- [13] R. Szeliski, “Computer Vision : Algorithms and Applications,” 2010.
- [14] M. Brown and D. Lowe, “Automatic Panoramic Stitching Using Invariant Features,” *International Journal on Computer Vision (IJCV)*, vol. 74, no. 1, pp. 59–73, 2007.
- [15] M. Hassaballah, A. A. Abdelmgeid, and H. A. Alshazly, *Image Features Detection , Description and Matching*.
- [16] M. Brown and D. G. Lowe, “Recognising panoramas,” *Proceedings of the {IEEE} International Conference on Computer Vision*, pp. 1218–1225, 2003.
- [17] S. Fountas, S. M. Pedersen, and S. Blackmore, “ICT in Precision Agriculture – diffusion of technology,” *Sciences-New York*, no. July 2015, pp. 1–15, 2005.
- [18] D. García-cervigón and J. José, “Estudio de Índices de vegetación a partir de imágenes aéreas tomadas desde UAS / RPAS y aplicaciones de estos a la agricultura de precisión .,” p. 78, 2015.
- [19] S. Aggarwal, “Principles of remote sensing,” *Satellite Remote Sensing and GIS Applications in Agricultural Meteorology*, pp. 23–38, 2004.
- [20] G. Butcher, “Tour of the Electromagnetic Spectrum,” *National Aeronautics and Space Administration*, p. 17, 2010.
- [21] W. Tomasi, I. Gloria, M. Hernández, I. Virgilio, and G. Pozo, *Sistemas de Comunicaciones Electrónicas*. Pearson Educación, cuarta ed., 2003.
- [22] B. Fontal, “El Espectro Electromagnético y sus Aplicaciones,” *Escuela de la ingeniería*, vol. 1, p. 24, 2005.
- [23] I. Press, R. J. Herrick, E. Chief, J. B. Anderson, P. M. Anderson, and P. Laplante, “Introduction To Dwdm Technology,”
- [24] J. Flammer, M. Mozaffarieh, and H. Bebie, “Basic Sciences in Ophthalmology,” *Antimicrobial Agents and Chemotherapy*, vol. 58, no. 12, pp. 7250–7257, 2013.
- [25] E. Falkner and D. Morgan, *Aerial Mapping Methods and Applications*. Boca Raton: Lewis Publishers, 2nd ed., 2002.
- [26] F. Group, *Image Sensors and Signal Processing for Digital Still Cameras*. 2006.
- [27] H. Eisenbeiß, *UAV Photogrammetry*. PhD thesis, Eidgenössische Technische Hochschule Zürich, 2009.
- [28] Dirección General de Aeronáutica Civil, *Dan 151: Operaciones de aeronaves pilotadas a distancia (RPAS) en asuntos de interés público, que se efectúen sobre áreas pobladas*.



2015.

- [29] J. C. McGlone, E. M. Mikhail, and J. S. Bethel, *Manual of Photogrammetry*. American Society for Photogrammetry and Remote Sensing (ASPRS), 5th ed ed., 2004.
- [30] R. Mlambo, I. H. Woodhouse, F. Gerard, and K. Anderson, “Structure from Motion ( SfM ) Photogrammetry with Drone Data : A Low Cost Method for Monitoring Greenhouse Gas Emissions from Forests in Developing Countries,” *Forests*, pp. 1 – 20, 2017.
- [31] E. Alegre and L. Fernández-Robles, “Capítulo 8: SIFT (Scale Invariant Feature Transform),” no. 1954, pp. 131–157.
- [32] T. Lindeberg, *Image Matching Using Generalized Scale-Space Interest Points*, vol. 52. Springer US, 2015.
- [33] J. Ramesh, K. Rangachar, and S. Brian, *Machine vision*. McGraw-Hill, 1995.
- [34] E. Dubrofsky, “Homography Estimation,” *Optical Engineering*, vol. 15, no. March, p. 977, 2009.
- [35] D. Kriegman, “Homography Estimation,” *Optical Engineering*, vol. 15, no. March, p. 977, 2007.
- [36] I. Markelic, “Tutorial on Homographies,” *Markelic.De*, pp. 1–11.
- [37] S.-d. Ponds and P. D. Criteria, “Chapter 5 Table of Contents,” no. March, pp. 57–64, 2000.
- [38] H. Anton and C. Rorres, *Elementary Linear Algebra, Applications Version*. tenth ed.
- [39] K. G. Derpanis, “Overview of the RANSAC Algorithm,” *Image Rochester NY*, vol. 4, no. 1, pp. 2–3, 2010.
- [40] D. P. Capel, “Image mosaicing and super-resolution,” *Springer Science & Business Media*, p. 263p., 2001.
- [41] V. Rankov, R. J. Locke, R. J. Edens, and P. R. Barber, “An algorithm for image stitching and blending,” vol. 5701, no. March, pp. 190–199, 2005.
- [42] N. Snavely, S. M. Seitz, and R. Szeliski, “Modeling the World from Internet Photo Collections,” no. January, 2007.
- [43] F. Chiabrando, E. Donadio, and F. Rinaudo, “SfM For Orthophoto Generation: A Winning Approach For Cultural Heritage Knowledge,”
- [44] W. K.pratt, *Digital Image Processing*, vol. 53. 2013.
- [45] J. L. Schonberger and J.-M. Frahm, “Structure-from-Motion Revisited,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4104–4113, 2016.

- [46] G. Yi, L. Jianxin, Q. Hangping, and W. Bo, "Survey of structure from motion," *Proceedings of 2014 International Conference on Cloud Computing and Internet of Things, CCIOT 2014*, no. Cctot, pp. 72–76, 2014.
- [47] M. V. Anjana and L. Sandhya, "Implementation and Comparison of Feature Detection Methods in Image Mosaicing," vol. 2017, pp. 7–11, 2017.
- [48] H. Joshi, "Image Mosaicing using Harris , SIFT Feature Detection Algorithm," *International Journal of Science, Engineering and Technology Research (IJSETR)*, vol. 2, no. 11, pp. 2078–2082, 2013.
- [49] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.
- [50] N. Snavely, "Scene Reconstruction and Visualization from Internet Photo Collections: A Survey," *IPSJ Transactions on Computer Vision and Applications*, vol. 3, pp. 44–66, 2011.
- [51] M. a. Fischler and R. C. Bolles, "Paradigm for Model," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [52] L. Assaely and H. Juárez, "Álgebra Lineal Numérica, Mínimos Cuadrados y Optimización," *Universidad Autónoma Metropolitana, Departamento de Matemáticas*.
- [53] K. Madsen, H. B. Nielsen, and O. Tingleff, "Methods for Non-Linear Least Squares Problems," *Infomatics and Mathematical Modeling*, vol. 2, pp. 1–30, 2004.
- [54] Y. Xu, J. Ou, H. He, X. Zhang, and J. Mills, "Mosaicking of Unmanned Aerial Vehicle imagery in the absence of camera poses," *Remote Sensing*, vol. 8, no. 3, 2016.
- [55] W. A. Kalender, *Computed Tomography, Fundamentals, System Technology, Image Quality, Applications*. Publicis, 2011.
- [56] "senseFly, 'Datasets for Education'. [En línea]. Disponible: [www.sensefly.com/education/datasets/](http://www.sensefly.com/education/datasets/). [Accedido: 11-Mar-2018]."