

Estimación de homografías en imágenes aéreas georeferenciadas para construcción de mosaicos.

Ing. Xavier Sierra Canto



Maestría en Ciencias Matemáticas.
Facultad de Matemáticas
Universidad Autónoma de Yucatán
2011

Declaración.

En cumplimiento de uno de los requisitos para la titulación en la Maestría en Matemáticas de la Universidad Autónoma de Yucatán, dirijo el presente documento como tesis de Maestría para obtener el grado de Maestro en Ciencias Matemáticas.

Declaro que esta tesis fue realizada enteramente por mi y describe mi propio trabajo de investigación con excepción de las partes que así se indiquen.

Ing. Xavier Sierra Canto
Mérida, Yucatán
México
17 de enero de 2012

Agradecimientos.

- A mi esposa Cris, compañera y amiga incondicional, por todas las noches que pasó escuchando acerca de fórmulas y algoritmos y las que, por motivos de estudio, no hubo oportunidad de platicar (aunque a la mañana siguiente si, no se salvaba), gracias por todo el apoyo.
- A mis hijos, Xavier y Ariadna, que son mi más grande fuente de inspiración y superación, por mantenerme ocupado todo el día pues eso me cargaba de energía para trabajar en la noche si no me quedaba dormido en su cama.
- Al Dr. Arturo Espinosa quien ha probado conmigo su capacidad de paciencia, prueba que superó con creces, gracias por todos los conocimientos transferidos y por hacer de este trabajo una lucha hombro con hombro, es un ejemplo que como profesor me llevo a mi vida profesional.
- Al Dr. Ricardo Legarda quien como coordinador del programa de posgrado me dio todo su apoyo y lo hizo extensible como director de este trabajo de tesis. También probó su capacidad de paciencia al tenerme como estudiante en procesamiento de imágenes.
- A mis padres, Aurora y Gilberto, quienes no solo me han dado la vida, sino que además se han encargado de velarla, apoyarla y enriquecerla con su cercanía.
- A mis hermanos, Aurora y Gilberto, quienes como personas exitosas me impulsan a esforzarme para seguir su ejemplo.
- A mis compañeros del posgrado, gracias por conocernos y hacernos amigos.
- A Dios, sin él nada.

Resumen.

En el presente trabajo se plantea y aborda la problemática de estimar las homografías entre imágenes aéreas haciendo uso de dispositivos de posicionamiento global y acelerómetros, montados en un vehículo no tripulado de bajo costo, en el cual se encuentra la cámara. Para ello se hace uso de matrices de rotación y de la fórmula de Rodrigues para determinar la rotación y traslación de la cámara entre una imagen y la siguiente. Con ayuda de la matriz de calibración de la cámara se hallan al menos cuatro correspondencias con las cuales se determina la homografía entre ellas. Con dicha homografía se presenta la metodología para estimar el área de traslape entre ellas, desechando las que tienen escasa o amplia área de traslape. Para la construcción del mosaico es necesario determinar de manera más robusta la homografía pero se hace uso de la primera estimación para acelerar dicho proceso en la búsqueda de correspondencias. Se adquirieron imágenes al norte de la ciudad de Mérida con lo que fue posible demostrar la obtención de la homografía haciendo uso de la posición global de la cámara y de la persona en tierra que controla el sistema de adquisición.

*Dile a la mañana
que se acerca mi sueño,
que lo que se espera
con paciencia se logra.*
Juan Luis Guerra

Índice general

| | |
|--|-------------|
| Declaración. | II |
| Agradecimientos. | III |
| Resumen. | IV |
| Lista de figuras | VIII |
| Lista de tablas | IX |
| Lista de Algoritmos | X |
| 1. Introducción. | 1 |
| 1.1. Metodología. | 2 |
| 1.2. Objetivos | 3 |
| 1.2.1. Objetivo General | 3 |
| 1.2.2. Objetivos Específicos | 3 |
| 1.3. Estructura de la Tesis. | 4 |
| 2. Fundamentos. | 5 |
| 2.1. Descripción del sistema de adquisición | 5 |
| 2.1.1. Módulo GPS | 9 |
| 2.1.2. Acelerómetro | 9 |
| 2.1.3. Sistema Picavet | 10 |
| 2.2. Los sistemas de posicionamiento global | 11 |
| 2.3. La rotación y movimiento de cuerpos rígidos | 14 |
| 2.4. El modelo de cámara | 18 |
| 2.5. Homografías y el algoritmo DLT | 20 |
| 3. Modelado del sistema de adquisición | 24 |
| 3.1. Cambio de marco de referencia | 24 |
| 3.2. Rotación y traslación de la cámara | 26 |
| 3.3. Análisis de estabilidad del modelo | 30 |
| 3.4. Primera estimación de la homografía | 32 |

| | | |
|-----------|---|-----------|
| 3.5. | Determinación del área de traslape | 34 |
| 3.5.1. | Intersección de segmentos | 35 |
| 3.5.2. | Área de un polígono irregular | 36 |
| 4. | Construcción de mosaicos. | 38 |
| 4.1. | Detección de esquinas | 38 |
| 4.2. | Correspondencias | 41 |
| 4.3. | Segunda estimación de la homografía | 42 |
| 4.4. | Registro geométrico | 47 |
| 4.5. | Proyección en un plano | 49 |
| 4.6. | Mezclado | 50 |
| 4.7. | Resumen | 50 |
| 5. | Trabajo experimental. | 52 |
| 5.1. | Primera estimación de la homografía | 52 |
| 5.2. | Conclusiones | 58 |
| A. | Filtro de Kalman | 62 |
| | Bibliografía | 66 |

Índice de figuras

| | |
|---|----|
| 1.1. Esquema de la construcción de mosaicos | 2 |
| 2.1. Esquema del sistema de adquisición. La cámara es elevada a través de un papalote con el eje Y siempre orientado hacia la persona en tierra | 6 |
| 2.2. Placa Arduino Mega utilizada para el control de disparo de la cámara y manejo del GPS y acelerómetro | 7 |
| 2.3. Cámara digital utilizada. Puede observarse que se manipuló para poder ser disparada por un dispositivo electrónico | 8 |
| 2.4. Sistema electrónico final. Se observa la cámara, la tarjeta Arduino y los componentes electrónicos usados para la captura y etiquetado de las imágenes | 8 |
| 2.5. Leadtek LR9952. GPS con antena integrada de 20 canales | 9 |
| 2.6. Tarjeta con acelerómetro LIS3LV02DQ de tres ejes con capacidad de medición de hasta $\pm 6g$ | 10 |
| 2.7. Sistema Picavet proporciona estabilidad a la cámara al tener un juego de poleas que la mantienen hacia abajo | 11 |
| 2.8. Sistema Picavet | 12 |
| 2.9. Dos elipses, una para ajustar una región y otra para todo el planeta. | 13 |
| 2.10. Coordenadas de un punto en unidades geométricas y cartesianas | 14 |
| 2.11. Movimiento de un cuerpo rígido | 17 |
| 2.12. Geometría del modelo de cámara pinhole. | 18 |
| 3.1. Cambio de marco de referencia: primero de $C_o \rightarrow C_p$ y luego de $C_p \rightarrow C_s$ | 25 |
| 3.2. Cuatro posiciones, P1 y P3 son posiciones en tierra mientras que P2 y P4 posiciones de la cámara en el aire | 26 |
| 3.3. Marcos de referencia de las cuatro posiciones | 27 |
| 3.4. Los ejes Y de ambas cámaras resultan alineados en la última rotación lo que define la rotación final en Z entre dos imágenes | 28 |
| 3.5. Diferentes vectores de aceleración. a) No existe aceleración en X y Y por lo que el vector es la gravedad y la cámara se encuentra ortogonal a la superficie terrestre. b) existe un balanceo sobre el eje Y lo que produce un vector de aceleración con componentes en los tres ejes. c) Existe un balanceo en el eje X lo que produce un vector de aceleración en los tres ejes. | 29 |

| | | |
|------|---|----|
| 3.6. | Valores del jacobiano alrededor de un hemisferio | 32 |
| 3.7. | Valores del jacobiano alrededor de una posición fija | 33 |
| 3.8. | Dos imágenes con una área de traslape | 35 |
| 4.1. | Proceso para la formación de un mosaico (Capel [2001]) | 39 |
| 4.2. | Ejemplo de estimación de homografías con RANSAC | 48 |
| 4.3. | Proyección en una superficie plana | 49 |
| 4.4. | Proyección en una superficie cilíndrica | 50 |
| 5.1. | Imágenes utilizadas para la calibración de la cámara | 53 |
| 5.2. | Secuencia de cuatro imágenes capturadas con su posición global al norte de la ciudad de Mérida | 54 |
| 5.3. | Los marcos locales de la cámara al momento de tomar la imagen para <i>aire2</i> y <i>aire3</i> | 55 |
| 5.4. | Las imágenes <i>aire2</i> y <i>aire3</i> transformadas al marco global usado como referencia | 56 |
| 5.5. | Rotación de la cámara si los marcos locales de <i>aire2</i> y <i>aire3</i> estuviesen en la misma posición | 57 |
| 5.6. | Dos imágenes con área de traslape suficiente y rotación de la cámara entre ambas | 57 |
| 5.7. | Aplicación de la matriz de homografía encontrada (prácticamente una rotación) a una de las imágenes de la secuencia | 59 |
| 5.8. | Superposición de dos imágenes (referencia y transformada) a manera de mosaico | 59 |

Índice de cuadros

| | |
|---|----|
| 3.1. Variación en X, Y y Z | 33 |
| 5.1. Coordenadas de posición de las cuatro imágenes | 53 |
| 5.2. Correspondencias para cuatro puntos del espacio y dos instantes de la cámara | 58 |

Lista de Algoritmos

| | | |
|------|---|----|
| 2.1. | <i>Algoritmo DLT.</i> | 22 |
| 3.1. | <i>Algoritmo para la primera estimación de la homografía.</i> | 34 |
| 3.2. | <i>Algoritmo para el área de traslape.</i> | 37 |
| 4.1. | <i>Algoritmo de detección de esquinas de Harris.</i> | 41 |
| 4.2. | <i>Algoritmo de obtención de correspondencias</i> | 42 |

Capítulo 1

Introducción.

La construcción de mosaicos ha representado una herramienta en diversas áreas en el campo de la visión computacional en conjunto con el uso de vehículos de bajo coste para la adquisición de imágenes aéreas. Jensen et al. [2009] describe el uso de múltiples vehículos aéreos no tripulados para la obtención de imágenes aéreas con el fin de realizar mediciones de dirección del viento así como fotogrametría en 3D. Usando puntos de control en tierra y dispositivos de posicionamiento en los vehículos aéreos, la reconstrucción tridimensional de la escena es posible. En Heikkilä and Pietikäinen [2005] se utiliza la construcción de mosaicos con el fin de ampliar el campo de visión en cámaras de vigilancia. La construcción del mosaico se realiza con imágenes de una secuencia de vídeo. En otros casos se han construido vehículos aéreos no tripulados para la adquisición de imágenes como muestra Oh and Green [2003] quienes utilizaron las imágenes obtenidas por su vehículo en la construcción de mosaicos y reconstrucciones tridimensionales de la escena. Botterill et al. [2010] realizaron recientemente la construcción de mosaicos en tiempo real con el objetivo de realizar los mosaicos en secuencias de vídeo obtenidas para ampliar el rango de visión de los pilotos aéreos. Inclusive, los dispositivos móviles son ahora partícipes de programas que construyen mosaicos para diversos fines. Xiong et al. [2009] describen la implementación de un programa para la construcción de mosaicos de alta definición en teléfonos móviles. Aun mas allá, Hannuksela et al. [2007] utilizan la construcción de mosaicos para escanear documentos con la cámara de un teléfono móvil usándola como cámara de vídeo.

En todos los casos anteriores la construcción del mosaico requiere del registro geométrico de las imágenes. Como se verá mas adelante, dicho proceso puede resultar

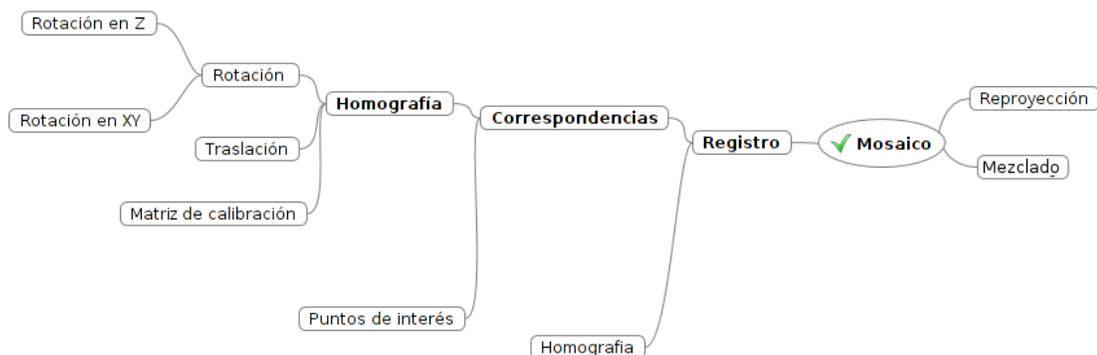


Figura 1.1: Esquema de la construcción de mosaicos

exhaustivo, computacionalmente hablando, pues requiere de la verificación de un gran número de puntos de interés con el objetivo de hallar correspondencias entre dos imágenes.

El presente trabajo describe precisamente una metodología que permite utilizar la información de posicionamiento global de un sistema de adquisición de imágenes aéreas de modo que puede simplificarse el procedimiento de búsqueda de correspondencias. El sistema de adquisición tiene características especiales que imponen restricciones a la metodología presentada en este trabajo y que se describirán con detalle mas adelante.

1.1. Metodología.

Las etapas para la estimación de la homografía entre imágenes aéreas georreferenciadas y la construcción de mosaicos se presentan en la figura 1.1. Esto determina la metodología y estructura del presente trabajo de tesis como se describe a continuación.

Como se indica en la figura 1.1 es necesario determinar la rotación y traslación de la cámara de una imagen a otra. En este caso se debe primero determinar la rotación sobre el eje Z y la rotación que pudo existir sobre el eje X o Y o ambos. Utilizando esta información y la información intrínseca de la cámara es posible obtener una primera estimación de la transformación lineal que mapea los píxeles de una imagen sobre el plano de otra imagen de referencia. Dicha transformación se conoce como *homografía* y una vez obtenida se utilizará para determinar las correspondencias entre imágenes de una forma más rápida y directa que los métodos tradicionales sin conocimiento de

dicha transformación. Además nos proporciona información del área de traslape entre imágenes de forma que los pasos subsecuentes en la construcción del mosaico se aplicarán únicamente entre imágenes que compartan una porción aceptable de la escena. Estas correspondencias servirán de base para obtener una segunda estimación de la homografía entre las imágenes con métodos robustos ante el ruido. Finalmente, se deben seguir los tres pasos básicos para la construcción del mosaico que son:

- 1 El registro geométrico usando la homografía entre imágenes.
- 2 La reproyección de los píxeles de las imagen en un único plano.
- 3 La unión o mezcla de imágenes.

1.2. Objetivos

1.2.1. Objetivo General

El objetivo general del trabajo es la obtención de homografías entre imágenes aéreas de modo que se puedan construir mosaicos al combinar múltiples vistas de una misma escena.

1.2.2. Objetivos Específicos

- Construir un sistema de adquisición de imágenes aéreas etiquetadas con la posición global de la cámara y la aceleración al momento de adquirir la imagen. Dicho sistema deberá ser no tripulado y contar con su propia fuente de energía.
- Modelar el sistema de adquisición para la estimación de la rotación y traslación de la cámara en cada imagen.
- Determinar una metodología para estimar el área de traslape ente imágenes contiguas antes de adquirir una imagen nueva de forma que el sistema de adquisición pueda determinar si es útil o no adquirir la siguiente imagen.
- Implementar algoritmos apropiados para la construcción de mosaicos en imágenes aéreas

1.3. Estructura de la Tesis.

De lo anterior, la estructura del presente trabajo resulta de la siguiente forma:

- El segundo capítulo presenta las bases necesarias para la adecuada comprensión del trabajo como son los sistemas de posicionamiento global, las matrices de rotación y el modelo de cámara y un algoritmo que se utilizará en la estimación de homografías: DLT, así como una descripción del sistema de adquisición.
- En el capítulo tres se encuentra la forma en que se modela el sistema para estimar la rotación y traslación de la cámara en cada imagen obtenida, la obtención del área de traslape y la obtención de la primera estimación de la homografía entre imágenes contiguas.
- El capítulo cuatro presenta los algoritmos utilizados en la construcción de los mosaicos, la detección de esquinas y puntos de interés, la forma en que se utiliza la primera estimación de la homografía para obtener correspondencias de una forma guiada y la estimación robusta de la homografía, la rectificación geométrica de las imágenes y la construcción del mosaico.
- El capítulo cinco presenta los resultados experimentales con las imágenes obtenidas del sistema.
- Finalmente, las conclusiones y el trabajo futuro se presenta en el capítulo seis.

Capítulo 2

Fundamentos.

En este capítulo se presenta los fundamentos teóricos y matemáticos necesarios para el desarrollo de la tesis. En la primera sección se presentan la descripción del sistema de adquisición que servirá para comprender las características físicas del sistema y las restricciones que impone. En la segunda parte se abordan los fundamentos de los sistemas de posicionamiento global, las operaciones necesarias para ir de un sistema coordinado a otro y las aplicaciones que se les ha dado en la construcción de mosaicos con imágenes tanto aéreas como sobre tierra. La segunda sección presenta los fundamentos de las matrices de rotación y el movimiento de cuerpo rígido, el grupo al que pertenecen y su representación en coordenadas homogéneas, finalmente se hace una revisión del modelo de cámara *pinhole* que servirá de base para la primera estimación de la homografía, así como el algoritmo DLT necesario para llevar a cabo dicha estimación.

2.1. Descripción del sistema de adquisición

La figura 2.1 representa la idea principal del sistema de adquisición de imágenes aéreas georeferenciadas. En una primera instancia se encuentra el vehículo aéreo no tripulado cuyo movimiento y dirección depende únicamente de las condiciones de viento existentes en el momento. Dicho vehículo cuenta con una cámara para la adquisición de las imágenes, la resolución y velocidad de adquisición depende del modelo comercial que se este utilizando en el sistema. Además cuenta con una dispositivo para la medición de la aceleración en los tres ejes coordinados que servirá para determinar la rotación de la cámara sobre los ejes X y Y y un dispositivo GPS que proporcionará la posición global

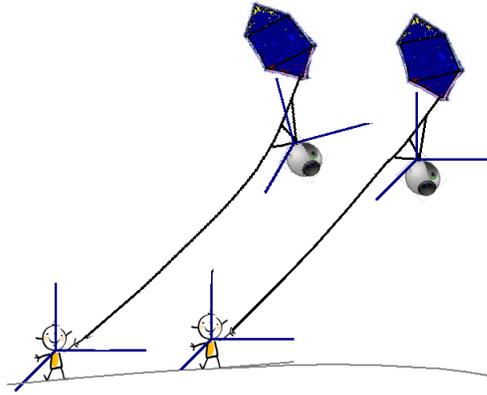


Figura 2.1: Esquema del sistema de adquisición. La cámara es elevada a través de un papalote con el eje Y siempre orientado hacia la persona en tierra

de la cámara al momento de adquirir una imagen. El vehículo se encuentra unido a una persona en tierra de forma que el eje Y de la cámara siempre se encuentra orientado hacia la dirección de dicha persona. En tierra, se cuenta con un GPS para establecer la posición del usuario al momento de adquirir una imagen, dato que será de gran utilidad en la determinación de la rotación en el eje Z de la cámara. De este modo, la persona puede moverse en cualquier dirección arrastrando al vehículo aéreo para cubrir la cantidad de área terrestre deseada en las imágenes.

El sistema cuenta con una tarjeta controladora basada en un micro controlador conocida como Arduino. Dicha tarjeta será la encargada de realizar cálculos y determinar la pertinencia de adquirir o no una imagen en determinado momento y posición. A continuación se presenta una descripción de los principales componentes del sistema.

Arduino es una plataforma de diseño de sistemas de control para quienes tienen poco o nulo conocimiento en electrónica. Una excelente referencia para el manejo y programación de estas tarjetas se encuentra en Margolis [2011]. La figura 2.1 muestra la placa de Arduino utilizada en este trabajo. El Arduino Mega es una placa microcontrolador basada en ATmega1280. Tiene 54 entradas/salidas digitales (de las cuales 14 proporcionan salida PWM), 16 entradas digitales, 4 UARTS (puertos serie por hardware), un cristal oscilador de 16MHz, conexión USB, entrada de corriente, conector ICSP y botón de reset. La placa puede trabajar con una alimentación externa de entre 6 a 20 voltios. El ATmega1280 tiene 128KB de memoria flash para almacenar código (4KB son usados

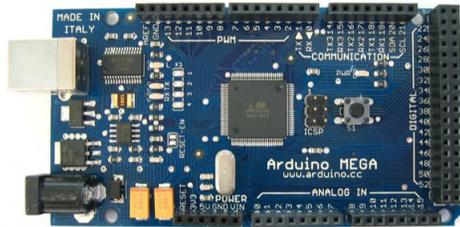


Figura 2.2: Placa Arduino Mega utilizada para el control de disparo de la cámara y manejo del GPS y acelerómetro

para el arranque del sistema *bootloader*), 8 KB de memoria SRAM y 4KB de EEPROM.

En este trabajo, se utilizaron un par de entradas y salidas digitales de modo que se pudiera accionar el disparador de la cámara y controlar el momento de descarga de la información de posición y aceleración almacenada en memoria RAM. Además, se utilizaron dos puertos de comunicación serial: un puerto RS-232 TTL para recibir los datos proveniente del GPS y otro puerto SPI para la recepción de la información del acelerómetro.

El ambiente de programación es multiplataforma por lo que se pueden utilizar sistemas operativos como windows o linux. El IDE se llama *Eclipse* y es de distribución gratuita lo que lo hace práctico y económico. La programación de la tarjeta se hace a través del puerto USB por lo que no se requiere de programadores externos para almacenar el programa.

La figura 2.1 es una fotografía de la cámara digital utilizada para el trabajo. Se trata de una BENQ DS-130 que fue desarmada para manipular el botón de disparo de modo que fuese posible accionarlo externamente a través de un relevador con contacto normalmente abierto. En la imagen se aprecia como se extrajeron dos cables y se probó con un botón externo dicho accionamiento.

La figura 2.1 muestra el sistema electrónico en su totalidad compuesto básicamente de: placa Arduino Mega, placa con chip GPS, tarjeta acelerómetro y chip MAX232 para comunicación serial con la computadora; relevador y batería de alimentación.



Figura 2.3: Cámara digital utilizada. Puede observarse que se manipuló para poder ser disparada por un dispositivo electrónico

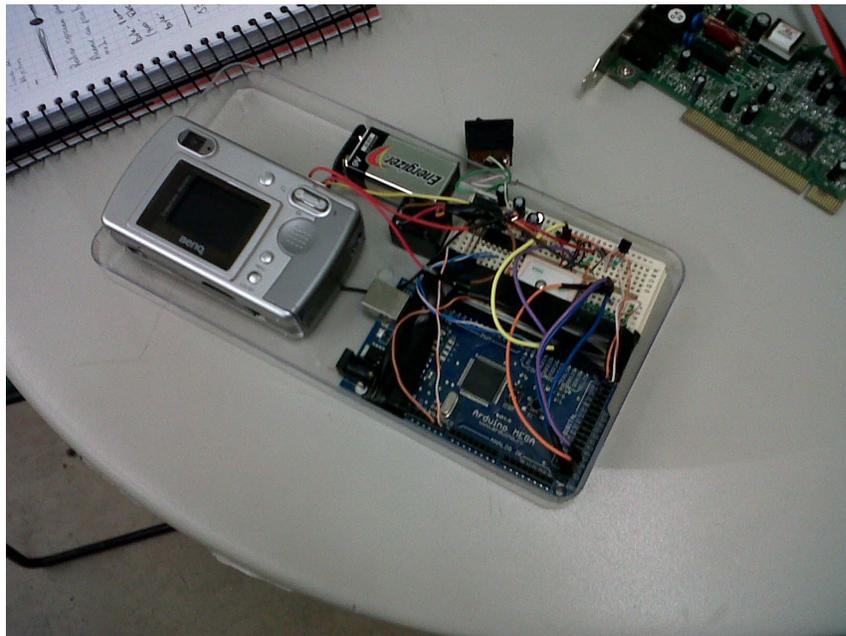


Figura 2.4: Sistema electrónico final. Se observa la cámara, la tarjeta Arduino y los componentes electrónicos usados para la captura y etiquetado de las imágenes



Figura 2.5: Leadtek LR9952. GPS con antena integrada de 20 canales

2.1.1. Módulo GPS

Como parte fundamental del sistema de adquisición se encuentra el receptor GPS utilizado para la adquisición de la posición global de la cámara al momento de capturar una imagen. Se utilizó el módulo GPS *Leadtek LR9952* el cual es de alta sensibilidad e incluye la antena lo que lo hace de dimensiones reducidas particularmente importante para esta aplicación.

Cuenta con 20 canales para la determinación de la posición y con una arquitectura basada en SiRFStarIII. Su precisión es de alrededor de 10 metros y utiliza el elipsoide WGS-84 y protocolo RS232 a nivel TTL para comunicación con la placa Arduino por lo que su conexión es directa con el puerto de comunicaciones de la placa. Los mensajes enviados son en base a la norma NMEA-0183 de la cual se utilizó la cadena GGA que se actualiza cada segundo.

La imagen 2.1.1 muestra una fotografía del circuito integrado utilizado.

2.1.2. Acelerómetro

Los acelerómetros son dispositivos para medir aceleración y vibración. Estos dispositivos convierten la aceleración de gravedad o de movimiento, en una señal eléctrica analógica proporcional a la fuerza aplicada al sistema, o mecanismo sometido a vibración o aceleración. Esta señal analógica indica en tiempo real, la aceleración instantánea del objeto sobre el cual el acelerómetro está montado. Los acelerómetros miden la aceleración en unidades g . Un g se define como la fuerza gravitacional de la tierra aplicada sobre un objeto o persona. Los acelerómetros son direccionales, esto quiere decir que sólo miden aceleración en un eje. Para monitorear aceleración en tres dimensiones, se emplea acelerómetros multi-ejes (ejes x,y,z), los cuales son ortogonales. Existen dos tipos



Figura 2.6: Tarjeta con acelerómetro LIS3LV02DQ de tres ejes con capacidad de medición de hasta $\pm 6g$

de acelerómetros a saber: pasivos y activos. Los acelerómetros pasivos envían la carga generada por el elemento sensor (puede ser un material piezoeléctrico), y debido a que esta señal es muy pequeña, estos acelerómetros requieren de un amplificador para incrementar la señal. Los acelerómetros activos incluyen circuitos internos para convertir la carga del acelerómetro a una señal de voltaje, pero requieren de una fuente constante de corriente para alimentar el circuito. Las opciones de salida eléctrica dependen del sistema utilizado con los acelerómetros. Las opciones analógicas comunes son voltaje, corriente, y frecuencia. Las opciones digitales son las señales paralelas y seriales. Otra opción es usar acelerómetros con una salida de cambio de estado de switches o alarmas.

Para obtener la aceleración en cada uno de los ejes debido a la inclinación de la cámara al momento de capturar una imagen, se utilizó una tarjeta basada en el acelerómetro *LIS3LV02DQ*. Dicho dispositivo es un circuito integrado micro-electromecánico capaz de medir aceleración en tres ejes con rango de $\pm 2g$ o $\pm 6g$ con salida digital con protocolo I2C o SPI. Ambos protocolos soportados por el puerto de comunicación de la placa Arduino. Para este trabajo se utilizó el protocolo SPI. La figura 2.1.2 muestra la tarjeta utilizada para la adquisición de la aceleración.

2.1.3. Sistema Picavet

El sistema Picavet es un arreglo de poleas y cuerdas que le da estabilidad al sistema de adquisición frente a movimientos bruscos provocados por ráfagas de viento. La figura 2.1.3 es una imagen del sistema picavet construido para este trabajo.

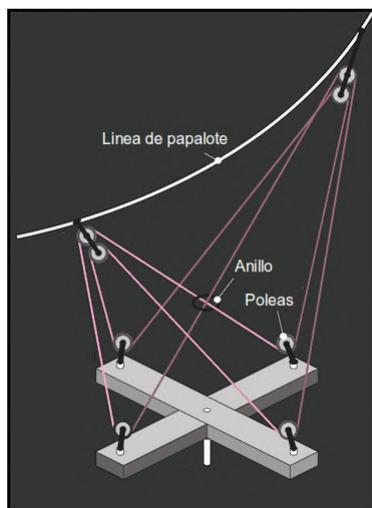


Figura 2.7: Sistema Picavet proporciona estabilidad a la cámara al tener un juego de poleas que la mantienen hacia abajo

El objetivo del sistema picavet es que la cámara esté el mayor tiempo posible mirando hacia abajo ayudando al sistema a tener menor rotación. El juego de poleas produce que, aun cuando el hilo sobre el que esté montado el sistema se haga mas vertical el sistema mantendrá posición horizontal.

2.2. Los sistemas de posicionamiento global

La creación de los sistemas de posicionamiento global (GPS por sus siglas en inglés) a través de señales obtenidas de satélites en órbita tiene sus inicios en los años cincuenta del milenio pasado. No fue hasta 1993 cuando el gobierno de los Estados Unidos declaró el sistema operacional al tener una constelación de satélites que asciende a 24 unidades (Xu [2007]). La forma en la que los satélites ayudan a determinar la posición de una persona en tierra es a través del tiempo que tarda en llegar una señal del satélite hasta el dispositivo de posicionamiento como se menciona en Kaplan et al. [2006]. Utilizando mínimo tres de los veinticuatro satélites es posible triangular la ubicación de la persona. La precisión de estos sistemas es de algunos metros para sistemas personales o de inclusive centímetros para sistemas industriales.

La Tierra tiene una forma muy complicada e irregular, por lo que es necesario un modelo que defina dicha forma sobre el cual un sistema coordenado esté basado. La



Figura 2.8: Sistema Picavet

ciencia de la *Geodesia* se encarga de determinar la forma y el tamaño de dicho modelo y de ahí, determinar la ubicación de puntos en la superficie terrestre. El modelo más simple que se aproxima a la forma y tamaño de la Tierra es una elipsoide biaxial que se construye al rotar una elipse sobre su eje mas corto. Dado que dicha elipse no se ajusta perfectamente, existen diferentes elipses en uso, algunas con la finalidad de ajustar la forma de todo el planeta y otras para ajustar la forma de una región en especial (ver Figura 2.9)(Agency [2010]).

En el caso de los sistemas de posicionamiento global utilizan una elipse denominada WGS-84 cuyo error se considera menor a 2 centímetros. Dicha elipse forma una esfera con radio mayor $a = 6,378,137 m$ en el ecuador y radio menor $b = 6,356,752.314 m$ en los polos.

La forma mas común de determinar la posición de un objeto es a través de dos ángulos: latitud (ϕ) y longitud (λ). Estos definen la posición en el globo, o de manera mas precisa, la posición sobre el elipsoide que se utilizó para aproximar la superficie terrestre por lo que es necesario saber que elipsoide es el que se esta utilizando para la medición. En la Figura 2.10 se puede observar la relación entre la longitud y latitud

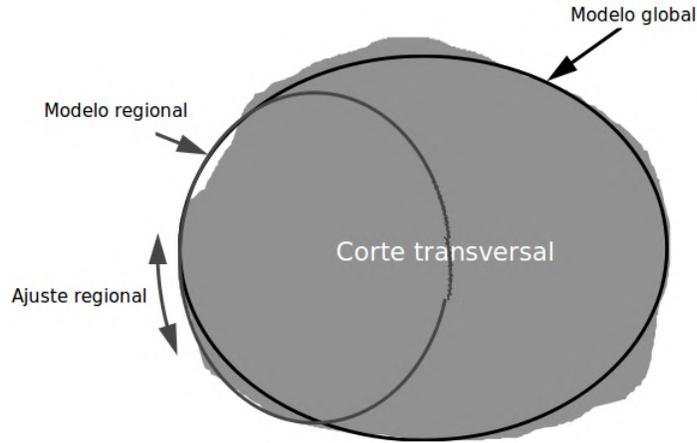


Figura 2.9: Dos elipses, una para ajustar una región y otra para todo el planeta.

con respecto a la elipsoide. Las líneas constantes en longitud que van de norte a sur son denominadas *meridianos*, mientras que las líneas constantes en latitud de este a oeste son denominadas *paralelas*. Una de las líneas de meridiano se toma como referencia al igual que una de las líneas paralelas (comúnmente el ecuador). De esta forma, la longitud de un punto es al ángulo entre el meridiano del punto y el de referencia. La latitud es el ángulo entre el plano ecuatorial del punto y el ecuador. De manera práctica la longitud se divide en hemisferio este y oeste determinando el ángulo de longitud entre 0-180 grados y la latitud se divide entre hemisferio norte y sur determinando el ángulo de latitud entre 0-90 grados. Ahora, dado que un punto pudiera estar por arriba o debajo del elipsoide es necesario un tercer parámetro llamado *peso*, el cual es simplemente una distancia entre la superficie del elipsoide y el punto, ya sea por encima o por debajo.

Es posible representar la posición de un punto en coordenadas cartesianas conociendo la longitud, latitud y peso. Como se muestra en la figura 2.10 es necesario alinear los ejes coordenados con respecto a la longitud y latitud cero y tomar como origen el geocentro de la elipsoide. La relación matemática entre un sistema coordenado y otro está dada por:

$$\begin{aligned} x &= (v + h)\cos\phi \cos\lambda \\ y &= (v + h)\cos\phi \sen\lambda \\ z &= ((1 - e^2)v + h)\sen\phi \end{aligned} \quad (2.1)$$

donde $e^2 = \frac{a^2 - b^2}{a^2}$ y $v = \frac{a}{\sqrt{1 - e^2 \sen^2\phi}}$

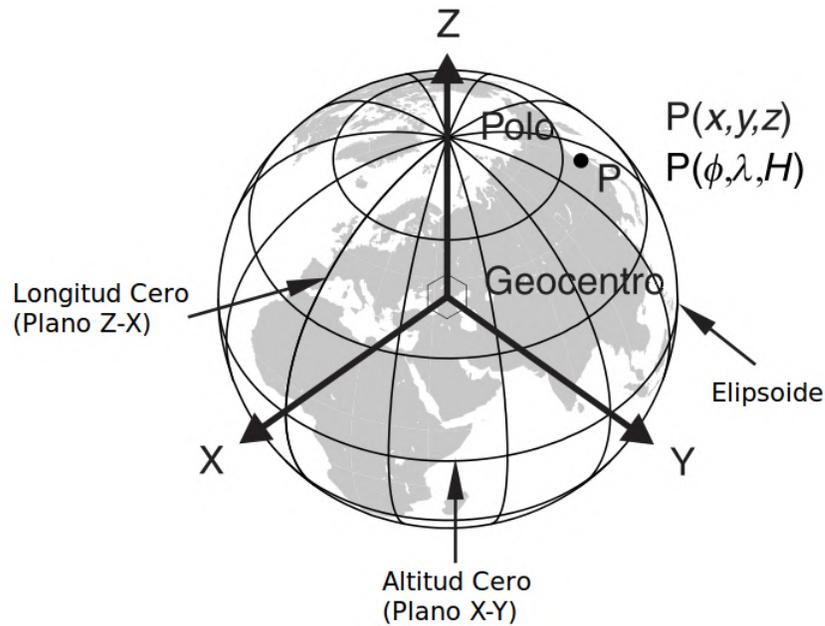


Figura 2.10: Coordenadas de un punto en unidades geométricas y cartesianas

Sin embargo, el GPS entrega información de latitud, longitud y altitud por lo que se desconoce el peso. Para estimar el valor del peso se puede utilizar la altitud (distancia del nivel medio del mar al punto) y la separación geoidal (diferencia entre el elipsoide WGS-84 y el nivel medio del mar) ambos parámetros conocidos a través del GPS.

$$h = \Lambda - g. \quad (2.2)$$

donde Λ es la altitud y g la separación geoidal. De esta forma, es posible obtener la posición de un punto P en coordenadas geométricas o cartesianas. En este caso, se utilizarán coordenadas cartesianas de manera que la posición de la cámara en un instante dado se puede representar como un vector $\mathbf{p} = [x, y, z]^T$.

2.3. La rotación y movimiento de cuerpos rígidos

En la mayor parte del trabajo realizado para la estimación de la homografía se trata con el movimiento de cuerpos rígidos. A continuación se hace una revisión teórica de las bases de dichos movimientos en su forma matricial, que pertenecen a un grupo en

especial. Únicamente se trata de establecer las características de dicho grupo que servirán en el desarrollo posterior del trabajo. Existe mucha mayor información acerca de dichas matrices y se pueden consultar fuentes que ahonden en el tema como Tapp [2000] y Curtis [1984].

Se empieza con una definición básica de la que se desarrollarán las demás definiciones y propiedades. Cabe mencionar que, aun cuando se puedan formar conjuntos y grupos con elementos en campos diferentes a los reales, todo lo que a continuación se establece es en el campo de los reales \mathbb{R}

Definición 2.3.1 Sea $\mathbf{M}_n(k)$ el conjunto de todas las matrices $n \times n$ con elementos $k \in \mathbb{R}$

Este conjunto de matrices no forman un grupo bajo el producto de matrices ya que algunas matrices no tienen inversa multiplicativa. Un ejemplo claro de esto es una matriz \mathbf{A} de ceros, no existe una matriz \mathbf{B} tal que $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. Las matrices que de hecho sí tienen un inverso multiplicativo forman un grupo especial denominado *Grupo Lineal General* (LG). La matriz \mathbf{B} que es el inverso de \mathbf{A} se denomina \mathbf{A}^{-1} .

Un subgrupo del LG es el denominado el Grupo Ortogonal (GO).

Definición 2.3.2 El grupo ortogonal sobre \mathbb{R} es denominado

$$O(n, \mathbb{R}) := \{\mathbf{A} \in LG(n, \mathbb{R}) \mid \mathbf{AA}^T = \mathbf{A}^T \mathbf{A} = \mathbf{I}\}$$

donde \mathbf{A}^T es la matriz transpuesta de \mathbf{A}

Es claro que $\mathbf{I} \in GO$ y que el inverso de \mathbf{A} es $\mathbf{A}^{-1} = \mathbf{A}^T$ por lo que con el producto de matrices forman un grupo. en lo subsecuente se denominará únicamente $O(n)$ pues como se mencionó anteriormente se trabaja sobre el campo de los reales.

Proposición 2.3.1 Si $\mathbf{A} \in O(n)$ entonces $|\det(\mathbf{A})| = 1$

Dado que $\mathbf{AA}^T = \mathbf{I}$, entonces $1 = \det(\mathbf{AA}^T) = \det(\mathbf{A})\det(\mathbf{A}^T) = |\det(\mathbf{A})|^2$

Entonces si $\mathbf{A} \in O(n)$ el $\det(\mathbf{A}) = \pm 1$. Existe entonces un conjunto de matrices en $O(n)$ a las que se les denomina Grupo Especial Ortogonal que es un subgrupo del anterior.

Definición 2.3.3 *El subgrupo de $O(n)$ llamado Grupo Especial Ortogonal es*

$$SO(n) = \{\mathbf{A} \in O(n) | \det(\mathbf{A}) = 1\}$$

Tradicionalmente este grupo de matrices son conocidas como *matrices de rotación*. La rotación de un cuerpo en movimiento constante puede definirse por la composición de rotaciones en cada instante, es decir,

$$\mathbf{R}(t_0, t_2) = \mathbf{R}(t_0, t_1)\mathbf{R}(t_1, t_2) \quad \forall t_0 < t_1 < t_2 \in \mathbb{R}$$

Es posible, sin embargo, definir una matriz de rotación con respecto al eje de rotación y el ángulo a través de la fórmula de Rodrigues . Ma et al. [2003] presenta una clara deducción de dicha fórmula que a continuación se resume y para ello se inicia con la definición de las matrices antisimétricas.

Definición 2.3.4 *Una matriz \mathbf{A}_n se dice antisimétrica si $\mathbf{A}^{-1} = -\mathbf{A}^T$*

Definición 2.3.5 *Dados dos vectores $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, su producto cruz es un tercer vector cuyas componentes están dadas por*

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix}$$

La importancia de las dos definiciones anteriores radica en que es posible representar el producto cruz de dos vectores en \mathbb{R}^3 como el producto punto entre una matriz y un vector. La matriz en cuestión es una matriz antisimétrica de la forma

$$\hat{\mathbf{u}} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}$$

Por ello, el producto de $\mathbf{u} \times \mathbf{v}$ se puede expresar como $\hat{\mathbf{u}}\mathbf{v}$. En general la fórmula de Rodrigues permite que dado un eje de rotación $\mathbf{w} = (w_x, w_y, w_z)$ unitario y un ángulo de rotación θ entonces,

$$\mathbf{R} = \mathbf{I} + \hat{\mathbf{w}} \sin \theta + \hat{\mathbf{w}}^2(1 - \cos \theta)$$

donde \mathbf{I} es la matriz identidad de 3×3 y $\hat{\mathbf{w}}$ es la matriz anti-simétrica

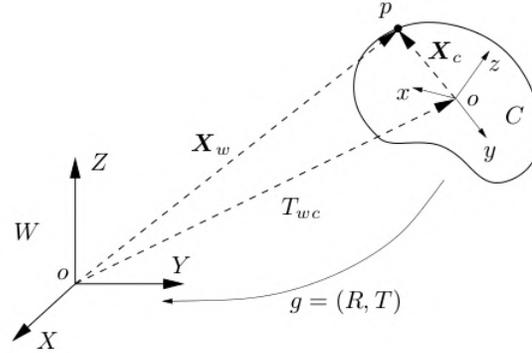


Figura 2.11: Movimiento de un cuerpo rígido

$$\begin{pmatrix} 0 & -w_z & w_y \\ w_z & 0 & -w_x \\ -w_y & w_x & 0 \end{pmatrix}$$

La matriz de rotación esta dada por:

$$\begin{bmatrix} \cos\theta + w_x^2(1 - \cos\theta) & w_x w_y(1 - \cos\theta) - w_z \sin\theta & w_y \sin\theta + w_x w_z(1 - \cos\theta) \\ w_z \sin\theta + w_x w_y(1 - \cos\theta) & \cos\theta + w_y^2(1 - \cos\theta) & -w_x \sin\theta + w_y w_z(1 - \cos\theta) \\ -w_y \sin\theta + w_x w_z(1 - \cos\theta) & w_x \sin\theta + w_y w_z(1 - \cos\theta) & \cos\theta + w_z^2(1 - \cos\theta) \end{bmatrix}$$

Hasta este punto se han definido algunas propiedades de las matrices de rotación y la forma de obtenerla conociendo el ángulo y el eje de rotación. Ahora se presentará la forma en la que se representa el movimiento de un cuerpo rígido (como la cámara) en su forma matricial de manera que incluya tanto la rotación como la traslación del cuerpo.

La figura 2.3 presenta una clara representación del movimiento de un cuerpo rígido con respecto a una marco de referencia global W . El punto \mathbf{p} tendrá coordenadas diferentes entre el marco global y el local (\mathbf{C}). La transformación que relaciona las coordenadas del punto entre los marcos esta definida por una rotación y una traslación. En el caso de la figura las coordenadas de \mathbf{p} con respecto a \mathbf{W} son \mathbf{x}_w y con respecto a \mathbf{C} son \mathbf{x}_c por lo tanto

$$\mathbf{x}_w = \mathbf{R}\mathbf{x}_c + \mathbf{T}$$

donde $\mathbf{R} \in SO(3)$ es la rotación relativa entre ambos marcos y \mathbf{T} la traslación.

Definición 2.3.6 *El conjunto de todas las posibles configuraciones de un cuerpo rígido se conoce como Transformaciones Euclidianas Especiales*

$$SE(3) = \{g = (\mathbf{R}, \mathbf{T}) \mid \mathbf{R} \in SO(3), \mathbf{T} \in \mathbb{R}^3\}$$

La representación matricial de dicho conjunto se logra gracias al uso de coordenadas homogéneas de forma que

$$SE(3) = \left\{ g = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0} & 1 \end{bmatrix} \mid \mathbf{R} \in SO(3), \mathbf{T} \in \mathbb{R}^3 \right\} \subset \mathbb{R}^{4 \times 4} \quad (2.3)$$

2.4. El modelo de cámara

A continuación se presenta el modelo de cámara *pinhole* que mapea puntos del espacio tridimensional a puntos en un plano proyectivo bidimensional. Dicho modelo, en conjunto con la rotación y traslación servir para llevar a cabo una primera estimación de la homografía existente entre dos imágenes dadas.

De acuerdo con Hartley and Zisserman [2004]: Sea el centro de proyección el origen de un sistema coordenado Euclideo y considerando el plano $z = f$, denominado el plano de la imagen o plano focal (figura 2.12), se tiene que un punto en el espacio \mathbf{X} con coordenadas $(X, Y, Z)^T$ se mapea al punto $\mathbf{x} = (fX/Z, fY/Z, f)^T$ en el plano de la imagen. Es decir,

$$(X, Y, Z)^T \mapsto (fX/Z, fY/Z, f)^T$$

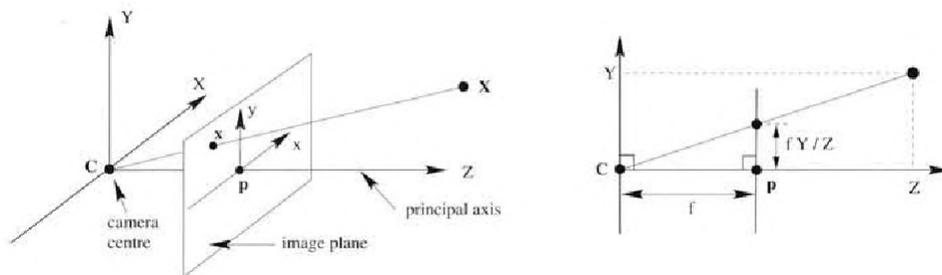


Figura 2.12: Geometría del modelo de cámara pinhole.

En coordenadas homogéneas, esta proyección es representada por un mapeo lineal entre sus coordenadas homogéneas. Particularmente puede representarse como un producto matricial,

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & & 0 \\ & f & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

De forma compacta puede decirse que

$$\begin{bmatrix} f & & 0 \\ & f & 0 \\ & & 1 & 0 \end{bmatrix} = \text{diag}(f, f, 1)[I|0] = P$$

donde P es la matriz de proyección de la cámara. Si se tiene un punto \mathbf{X} en el espacio tridimensional, se proyecta a un punto \mathbf{x} en la imagen de la siguiente forma:

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

Esta aproximación supone que las coordenadas de un punto en la imagen son con respecto a un marco cartesiano cuyo origen coincide con el centro de proyección de la cámara. En la práctica esto no es así pues las coordenadas de un punto en la imagen están en relación a una de las esquinas de la imagen. Esta diferencia se puede compensar de la siguiente forma:

$$(X, Y, Z)^T \mapsto (fX/Z + px, fY/Z + py, f)^T$$

donde $(px, py)^T$ son las coordenadas del punto principal. En coordenadas homogéneas se tiene que

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} fX + Zpx \\ fY + Zpy \\ Z \end{bmatrix} = \begin{bmatrix} f & px & 0 \\ & f & py & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Si hacemos

$$K = \begin{bmatrix} f & px & 0 \\ & f & py & 0 \\ & & 1 & 0 \end{bmatrix}$$

entonces

$$\mathbf{x} = K\mathbf{X}$$

donde K es la matriz de calibración de la cámara .

El ultimo aspecto a considerar es que las coordenadas de un punto en el espacio generalmente están con referencia a un marco diferente al de los puntos en las imágenes, dicho marco es conocido como *marco global*. La relación entre el marco global y el marco local de la imagen esta definida por una rotación y traslación como se vio en la sección anterior. Es decir, sean las coordenadas de un punto en el espacio \mathbf{X}_w con respecto al marco global (de ahí el subíndice w) y \mathbf{X}_c las coordenadas del mismo punto con respecto al marco local de la cámara, entonces

$$\mathbf{X}_c = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \mathbf{X}_w$$

Esto permite completar el modelo de la cámara de forma que

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & px \\ 0 & f & py \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{KDGX} \quad (2.4)$$

La matriz \mathbf{K} es conocida como la matriz de parámetros intrínsecos de la cámara, mientras que la matriz \mathbf{G} es la de parámetros intrínsecos.

Como se vera en el capitulo siguiente, este modelo se utilizará para la primera estimación de la homografía entre dos imágenes del sistema.

2.5. Homografías y el algoritmo DLT

A continuación se presenta la definición de homografía y el algoritmo DLT de acuerdo con Hartley and Zisserman [2004]:

Definición 2.5.1 *Una homografía o transformación proyectiva, es un mapeo invertible h de \mathbb{P}^2 a si mismo tal que tres puntos \mathbf{x}_1 , \mathbf{x}_2 y \mathbf{x}_3 pertenecen a la misma línea si y solo si $h(\mathbf{x}_1)$, $h(\mathbf{x}_2)$ y $h(\mathbf{x}_3)$ también.*

La definición anterior se establece bajo una representación geométrica de incidencia en un línea. Se puede realizar una definición de homografía en un plano más algebraico con ayuda del siguiente teorema.

Teorema 2.5.1 *Un mapeo $h : \mathbb{P}^2 \rightarrow \mathbb{P}^2$ es una transformación proyectiva si y solo si existe una matriz \mathbf{H} no singular de 3×3 tal que para cada punto en \mathbb{P}^2 representado por un vector \mathbf{x} es cierto que $h(\mathbf{x}) = \mathbf{H}\mathbf{x}$.*

Por esto, se puede dar una definición de homografía en términos algebraicos de la siguiente forma:

Definición 2.5.2 *Una homografía es una transformación lineal sobre vectores homogéneos representada en un matriz no-singular de 3×3 :*

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (2.5)$$

Esta relación es válida si la escena es plana o si el desplazamiento de la cámara es pequeño. En este caso dada la altura a la que se encuentra el sistema de adquisición y que la escena no contiene objetos elevados como edificios, cerros o montañas, se puede considerar la imagen como plana. Si se conocen puntos correspondientes entre ambas imágenes la homografía define la siguiente relación:

$$\mathbf{x}' = \mathbf{H}\mathbf{x} \quad (2.6)$$

donde \mathbf{H} es una matriz de 3×3 .

Los puntos son representados en coordenadas homogéneas por lo que es correcto hacer énfasis en que dicha igualdad se cumple bajo un factor de escala $k \neq 0$ por lo que se tiene $\mathbf{x}' = k\mathbf{H}\mathbf{x}$, condición que se omite en el siguiente desarrollo pero que es importante tomar en cuenta al final de la estimación pues será necesario dividir la matriz final entre este valor de escala.

De la ecuación 2.6 es claro que $\mathbf{x}' \times \mathbf{H}\mathbf{x} = \mathbf{0}$ lo que nos conduce a un sistema lineal.

Usado h^{jT} para denominar a la j -ésima fila de la matriz \mathbf{H} tenemos

$$\mathbf{H}\mathbf{x}_i = \begin{pmatrix} \mathbf{h}^{1T}\mathbf{x}_i \\ \mathbf{h}^{2T}\mathbf{x}_i \\ \mathbf{h}^{3T}\mathbf{x}_i \end{pmatrix}.$$

Si $\mathbf{x}'_i = (x'_i, y'_i, w'_i)$ el producto cruz puede expresarse explícitamente como:

$$\mathbf{x}'_i \times \mathbf{H}\mathbf{x}_i = \begin{pmatrix} y'_i \mathbf{h}^{3T} \mathbf{x}_i - w'_i \mathbf{h}^{2T} \mathbf{x}_i \\ w'_i \mathbf{h}^{1T} \mathbf{x}_i - x'_i \mathbf{h}^{3T} \mathbf{x}_i \\ x'_i \mathbf{h}^{2T} \mathbf{x}_i - y'_i \mathbf{h}^{1T} \mathbf{x}_i \end{pmatrix}.$$

Dado que $\mathbf{h}^{jT} \mathbf{x}_i = \mathbf{x}'_i \mathbf{h}^j$ para $j = 1, 2, 3$ lo cual nos da como resultado tres ecuaciones que se pueden escribir de la siguiente forma:

$$\begin{bmatrix} \mathbf{0}^T & -w'_i \mathbf{x}'_i{}^T & y'_i \mathbf{x}'_i{}^T \\ w'_i \mathbf{x}'_i{}^T & \mathbf{0}^T & -x'_i \mathbf{x}'_i{}^T \\ -y'_i \mathbf{x}'_i{}^T & x'_i \mathbf{x}'_i{}^T & \mathbf{0}^T \end{bmatrix} \begin{pmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{pmatrix} = \mathbf{0} \quad (2.7)$$

La ecuación 2.7 es lineal con respecto a los parámetros de \mathbf{H} y aunque contiene tres ecuaciones, únicamente las dos primeras son linealmente independientes por lo que se puede eliminar la última quedando

$$\begin{bmatrix} \mathbf{0}^T & -w'_i \mathbf{x}'_i{}^T & y'_i \mathbf{x}'_i{}^T \\ w'_i \mathbf{x}'_i{}^T & \mathbf{0}^T & -x'_i \mathbf{x}'_i{}^T \end{bmatrix} \begin{pmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{pmatrix} = \mathbf{0} \quad (2.8)$$

Cada par de correspondencias produce dos ecuaciones para resolver \mathbf{H} . El sistema en la ecuación 2.8 es válido para cualquier representación homogénea de las coordenadas de píxeles aunque usualmente se hace $w'_i = 1$. Dicho sistema puede verse como $\mathbf{A}\mathbf{h} = \mathbf{0}$ siendo \mathbf{A} una matriz de $2n \times 9$, n el número de correspondencias. La solución de dicho sistema se puede obtener a través de la descomposición de valores singulares de la matriz \mathbf{A} . Si $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, siendo \mathbf{D} diagonal con valores positivos en orden descendiente, entonces \mathbf{h} es la última columna de \mathbf{V} . El algoritmo 2.1 presenta la secuencia de pasos para llevarlo a cabo.

Algoritmo 2.1: *Algoritmo DLT.*

1. Para cada correspondencia $x_i \leftrightarrow x'_i$ determinar la matriz como muestra la ecuación 2.8.
2. Agrupar dichas matrices para formar una matriz \mathbf{A} de $2n \times 9$ donde n es el número de correspondencias.
3. Obtener la descomposición de valores singulares de la matriz \mathbf{A}
4. Si $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, siendo \mathbf{D} diagonal con valores positivos en orden descendiente, entonces \mathbf{h} es la última columna de \mathbf{V} .
5. Obtener \mathbf{H} de \mathbf{h} .

Esta solución es sencilla de calcular con 4 correspondencias ya que la matriz \mathbf{A} será de 8×9 , sin embargo el resultado puede no ser el esperado ya que es necesario realizar

una normalización previa de los datos para hacerlos invariantes a transformaciones de similitud y geométricas. El proceso consiste en cuatro pasos:

- Centrar al origen las mediciones al multiplicar por $T = \begin{pmatrix} 1 & 0 & -\mu_x \\ 0 & 1 & -\mu_y \\ 0 & 0 & 1 \end{pmatrix}$ donde (μ_x, μ_y) son la media de las mediciones. Hacer lo mismo para la otra imagen con una matriz T' .
- Escalar el resultado anterior para que tengan un promedio de $\sqrt{2}$ multiplicando cada medición por la matriz $S = \begin{pmatrix} \frac{\sqrt{2}}{d} & 0 & 0 \\ 0 & \frac{\sqrt{2}}{d} & 0 \\ 0 & 0 & 1 \end{pmatrix}$ donde d es la media de las mediciones después del paso anterior.
- Aplicar DLT con los nuevos datos para hallar una primera \hat{H} .
- El resultado des-normalizado se obtiene haciendo $H = T^T S \hat{H} S' T'$.

Capítulo 3

Modelado del sistema de adquisición

3.1. Cambio de marco de referencia

El primer paso para determinar los parámetros extrínsecos de la cámara de acuerdo con la ecuación 2.4 es definir un marco de referencia global para todas las imágenes adquiridas por el sistema. En Slama [1966] se sugiere que para este tipo de aplicaciones el origen de dicho marco de referencia este cercano al área de análisis de forma que los dígitos de cada número de coordenada sean mínimos, además, el eje Z sea perpendicular a la superficie terrestre y el eje Y orientado hacia el norte. Cabe recordar que el sistema de posicionamiento global tiene un marco de referencia con origen en el centro de la tierra, el eje Z hacia los polos y el eje X apuntando hacia el meridiano de Greenwich.

La figura 3.1 presenta la secuencia de transformaciones que deben realizarse para cambiar el marco de referencia. Sea $\mathbf{p} = [p_x, p_y, p_z]^T$ las coordenadas de una posición en tierra dada por el GPS con respecto al marco de referencia inicial C_o con origen en el centro de la tierra. Los ejes de dicho marco de referencia son X_o , Y_o y Z_o y se desea rotar dicho marco de forma que Z_o sea perpendicular a la superficie terrestre. Usando la fórmula de Rodrigues, la matriz de rotación está dada por

$$\mathbf{R}_{\mathbf{p}} = \mathbf{I} + \hat{\mathbf{w}}_p \sin \theta_p + \hat{\mathbf{w}}_p^2 (1 - \cos \theta_p) \quad (3.1)$$

donde $\hat{\mathbf{w}}_p$ es la matriz antisimétrica del vector $\mathbf{w}_{\mathbf{p}} = \frac{Z_o \times \mathbf{p}}{\|Z_o \times \mathbf{p}\|}$, $\theta_p = \arccos \frac{Z_o \cdot \mathbf{p}^T}{\|Z_o\| \|\mathbf{p}\|}$ y $Z_o = [0, 0, 1]^T$.

El resultado de dicha rotación es un marco de referencia con ejes X_p , Y_p y Z_p . Posteriormente se requiere de rotar sobre Z_p de modo que el eje Y_p este orientado al

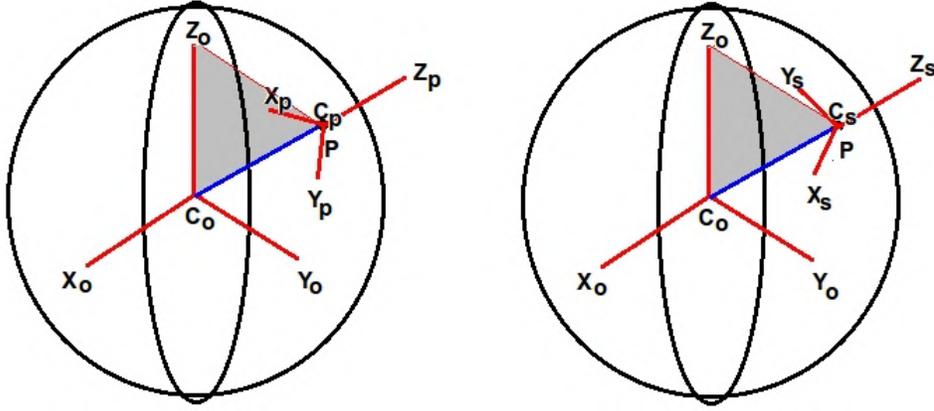


Figura 3.1: Cambio de marco de referencia: primero de $C_o \rightarrow C_p$ y luego de $C_p \rightarrow C_s$

norte. Dado que es una rotación sobre uno de los ejes coordenados la matriz esta definida como

$$\mathbf{R}_s = \begin{bmatrix} \cos \theta_s & -\sin \theta_s & 0 \\ \sin \theta_s & \cos \theta_s & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.2)$$

donde $\theta_s = 90 + longitud$.

La razón de este ángulo de rotación se encuentra en que el eje Y_c tiene una orientación que depende de la longitud a la que se encuentra el punto \mathbf{p} al momento de la transformación.

Finalmente, la traslación de un marco de referencia al otro esta determinada por las coordenadas del punto \mathbf{p} . De esta forma, la transformación puede representarse de forma matricial con el uso de coordenadas homogeneas siguiendo la ecuación 2.3.

$$\mathbf{G}_s = \begin{bmatrix} \mathbf{R}_T & \mathbf{T}_T \\ 0 & 1 \end{bmatrix} \quad (3.3)$$

donde $\mathbf{R}_T = \mathbf{R}_p \mathbf{R}_s$ y $\mathbf{T}_T = [p_x, p_y, p_z]^T$.

Cada nueva posición obtenida por el GPS se deberá cambiar a este marco de referencia.

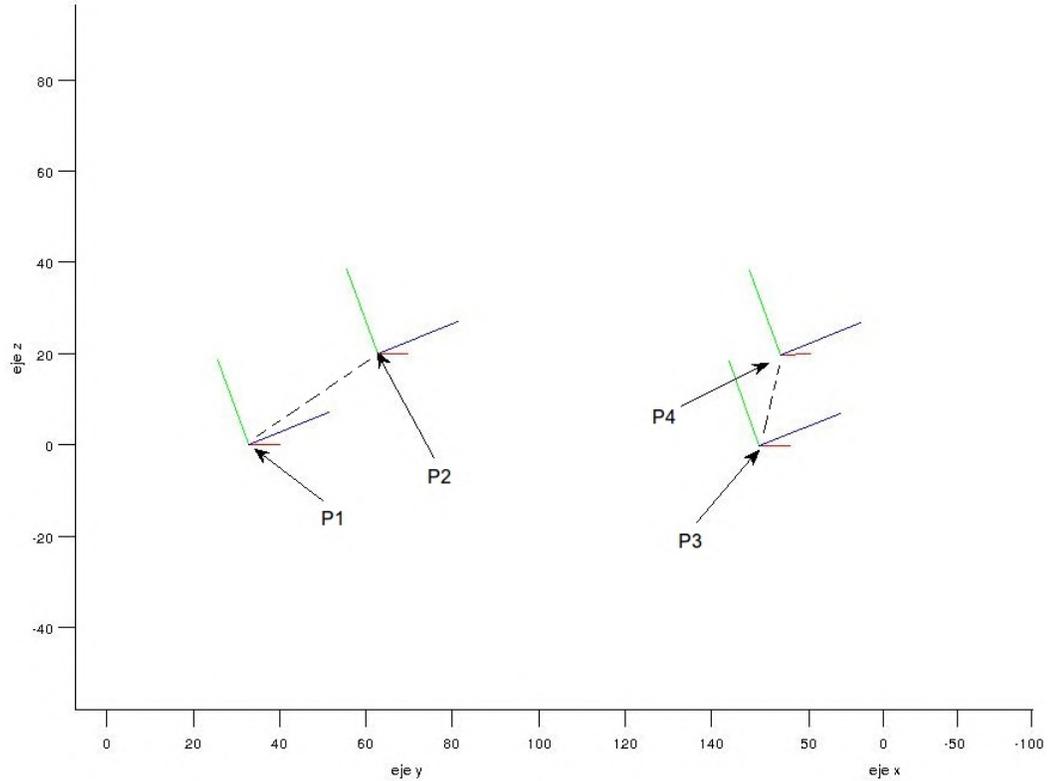


Figura 3.2: Cuatro posiciones, P1 y P3 son posiciones en tierra mientras que P2 y P4 posiciones de la cámara en el aire

3.2. Rotación y traslación de la cámara

A continuación es necesario resolver la rotación y traslación de la cámara entre una imagen y otra. La figura 3.2 muestra cuatro posiciones P obtenidas a través del GPS y que han sido transformadas a una marco de referencia local. La imagen hace referencia a tres vectores orientados en los ejes formando un sistema cartesiano por cada punto. El color azul correspondería al eje Z, el verde al eje Y y el rojo al X, la línea punteada de color negro es la que une a la persona con el equipo de adquisición.

Dado que una de las restricciones del sistema es que el eje Y de la cámara este orientado hacia la posición de la persona en tierra, la figura 3.3 es el resultado de rotar el eje Y hacia la posición en tierra. En la imagen el eje Z no se aprecia (color azul) ya que se realizó una rotación de la imagen para que dicho eje quedara alienado con la vista, es decir, se puede apreciar el plano X-Y, así, se puede observar como el eje Y (color

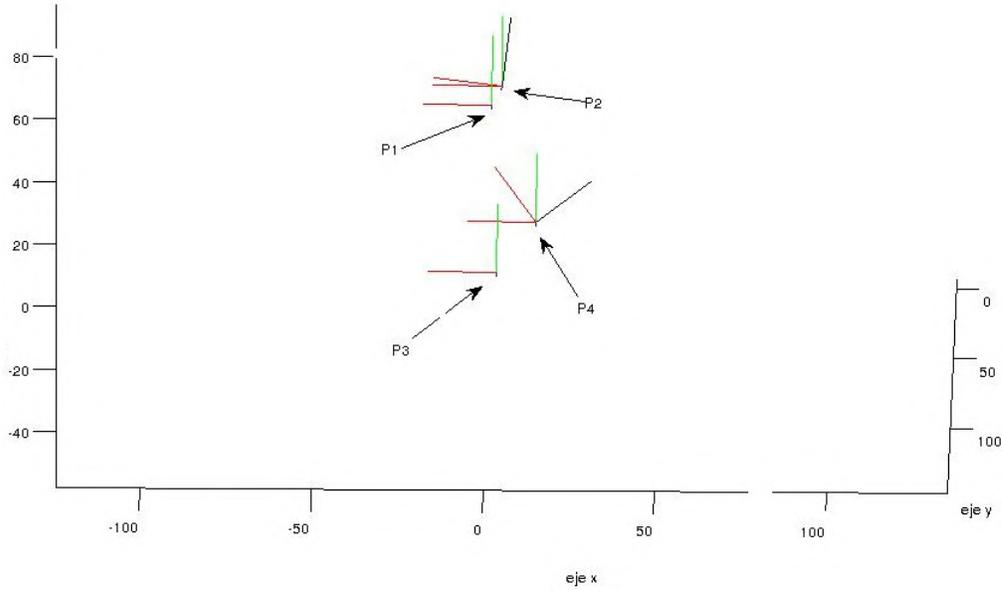


Figura 3.3: Marcos de referencia de las cuatro posiciones

verde) realiza una rotación sobre el eje Z para orientarse (color negro) hacia la posición en tierra.

Para lograr lo anterior se hace rotar el eje Y hacia la proyección en XY de la línea que los une:

Sea $\mathbf{P}_1 = [x_1, y_1, z_1]^T$ y $\mathbf{P}_2 = [x_2, y_2, z_2]^T$ las posiciones en tierra y aire que proporciona el GPS y $Y_{s1} = [a, b, c]^T$ el eje Y_s de la posición en el aire P_2 , la rotación se puede determinar usando nuevamente la fórmula de Rodrigues:

$$R_{y_1} = I + \hat{w} \sin \theta + \hat{w}^2 (1 - \cos \theta)$$

donde

$$\mathbf{w} = \begin{bmatrix} x_2 - x_1 \\ y_2 - y_1 \\ 0 \end{bmatrix} \times \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

y \hat{w} es la matriz antisimétrica formada con w . El ángulo de rotación se obtiene del producto punto de ambos como se vio anteriormente.

El mismo procedimiento se realiza para los puntos P_2 y P_3 con lo que se obtiene R_{y_2} . con ello, se tiene una representación de los ejes de la cámara tal como se da en el sistema

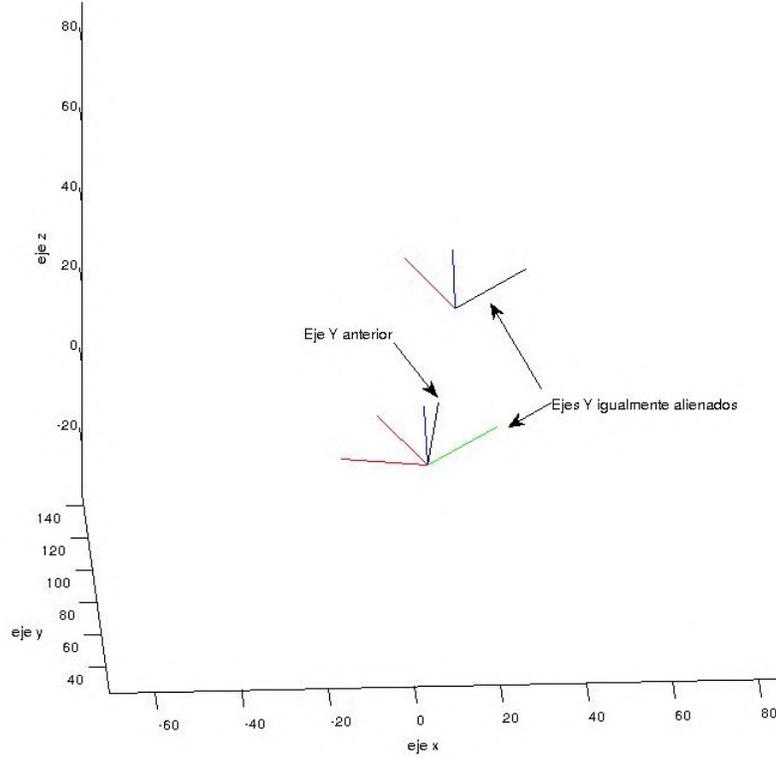


Figura 3.4: Los ejes Y de ambas cámaras resultan alineados en la última rotación lo que define la rotación final en Z entre dos imágenes

físico. Luego entonces, para determinar finalmente la rotación en Z de la cámara en dos instantes dados únicamente hay que determinar la rotación necesaria para que los ejes Y resulten iguales. La matriz está definida de la siguiente forma:

$$R_f = \begin{bmatrix} \cos\theta_f & -\text{sen}\theta_f & 0 \\ \text{sen}\theta_f & \cos\theta_f & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \theta_f = \text{arc Cos} \left(\frac{V_1 \cdot V_2}{|V_1| |V_2|} \right)$$

donde

$$V_1 = R_{y1} G_s \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad V_2 = R_{y2} G_s \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

La figura 3.4 muestra el resultado de aplicar la rotación de forma que ambos ejes Y se encuentran igualmente alineados por lo que se obtiene la rotación en Z ocurrida entre dos imágenes.

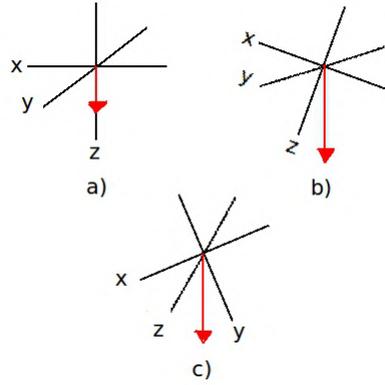


Figura 3.5: Diferentes vectores de aceleración. a) No existe aceleración en X y Y por lo que el vector es la gravedad y la cámara se encuentra ortogonal a la superficie terrestre. b) existe un balanceo sobre el eje Y lo que produce un vector de aceleración con componentes en los tres ejes. c) Existe un balanceo en el eje X lo que produce un vector de aceleración en los tres ejes.

Una vez obtenida la rotación con respecto al eje Z se requiere de estimar la rotación en los ejes X y Y. Para ello se cuenta con la información que proporciona un acelerómetro de 3 ejes. Dicho dispositivo nos entrega un vector de aceleración en cada eje $v = (a_x, a_y, a_z)^T$. La figura 3.5 muestra el vector que se forma al rotar la cámara sobre el eje X y Y. El inciso a) presenta un vector completamente orientado con el eje Z cuando el sistema se encuentra sin rotación sobre los demás ejes, el inciso b) muestra el resultado de una rotación en el eje Y y el inciso c) en el eje X. Esta representación nos lleva a definir la forma en que es necesario rotar un vector (inciso b y c) para alinearse al eje Z como se muestra en la figura.

Sea $\mathbf{a} = [(a_x, a_y, a_z)^T]$ el vector de aceleración y $\mathbf{a}_n = [0, 0, 1]^T$ el vector que apunta hacia abajo y hacia donde queremos rotar. Nuevamente por Rodrigues,

$$\mathbf{R}_{xy} = I + \hat{w} \sin \theta + \hat{w}^2 (1 - \cos \theta)$$

donde

$$w = \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

y \hat{w} es la matriz antisimétrica de w .

De esta forma se determina la rotación con respecto a los ejes X y Y. Como se vio anteriormente el producto de esta matriz con la matriz de rotación en el eje Z nos dará como resultado la matriz de rotación en los tres ejes.

La rotación final de la cámara es entonces:

$$\mathbf{R}_T = \mathbf{R}_z \mathbf{R}_{xy} \quad (3.4)$$

Para completar la matriz que define el movimiento de la cámara de un instante a otro se debe de calcular la traslación.

$$T_T = \begin{bmatrix} P_{x2} - P_{x4} \\ P_{y2} - P_{y4} \\ P_{z2} - P_{z4} \end{bmatrix}$$

Con ello la transformación entonces que define el movimiento de la cámara es :

$$G_f = \begin{bmatrix} R_{T(3x3)} & T_{T(3x1)} \\ 0_{(1x3)} & 1 \end{bmatrix} \quad (3.5)$$

3.3. Análisis de estabilidad del modelo

El objetivo de llevar a cabo el análisis de estabilidad del modelo es conocer la dimensión del ruido en el sistema de adquisición y la forma en que éste afecta al modelo. Para realizar el análisis de sensibilidad se utilizaron las derivadas parciales como razón de cambio de los parámetros de salida con respecto a los de entrada (Saltelli et al. [2004]). Dado que el proceso de transformación de un marco de referencia a otro es lineal, la sensibilidad del modelo se propaga en cada uno de los pasos descritos anteriormente, por ello, es necesario únicamente determinar la sensibilidad de los parámetros de salida (x, y ,z) ante perturbaciones en los parámetros de entrada (latitud, longitud y peso) al momento de transformar el vector de posición de coordenadas geodésicas a cartesianas.

Recordando las ecuaciones para transformar de coordenadas geodésicas a cartesianas es:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \left(\frac{a}{\sqrt{1-e^2 (\sin(\phi))^2}} + H \right) \cos(\phi) \cos(\lambda) \\ \left(\frac{a}{\sqrt{1-e^2 (\sin(\phi))^2}} + H \right) \cos(\phi) \sin(\lambda) \\ \left(\frac{(1-e^2)a}{\sqrt{1-e^2 (\sin(\phi))^2}} + H \right) \sin(\phi) \end{bmatrix}$$

donde $e^2 = 0.00669437999014$ y $a = 6,378,137$.

Las derivadas parciales pueden representarse en la matriz jacobiana de dicha transformación. La ecuación 3.6 presenta el resultado, donde la primera fila de la matriz son las derivadas parciales de x con respecto a latitud, longitud y peso respectivamente, la segunda fila son las derivadas de y y la última las de z .

$$\begin{bmatrix} \frac{ae^2 \sin(\phi) (\cos(\phi))^2 \cos(\lambda)}{(1-e^2 (\sin(\phi))^2)^{3/2}} - \left(\frac{a}{\sqrt{1-e^2 (\sin(\phi))^2}} + H \right) \sin(\phi) \cos(\lambda) & - \left(\frac{a}{\sqrt{1-e^2 (\sin(\phi))^2}} + H \right) \cos(\phi) \sin(\lambda) & \cos(\phi) \cos(\lambda) \\ \frac{ae^2 \sin(\phi) (\cos(\phi))^2 \sin(\lambda)}{(1-e^2 (\sin(\phi))^2)^{3/2}} - \left(\frac{a}{\sqrt{1-e^2 (\sin(\phi))^2}} + H \right) \sin(\phi) \sin(\lambda) & \left(\frac{a}{\sqrt{1-e^2 (\sin(\phi))^2}} + H \right) \cos(\phi) \cos(\lambda) & \cos(\phi) \sin(\lambda) \\ \frac{(1-e^2)ae^2 (\sin(\phi))^2 \cos(\phi)}{(1-e^2 (\sin(\phi))^2)^{3/2}} + \left(\frac{(1-e^2)a}{\sqrt{1-e^2 (\sin(\phi))^2}} + H \right) \cos(\phi) & 0 & \sin(\phi) \end{bmatrix} \quad (3.6)$$

Para efectos del análisis de sensibilidad se procede a simplificar el modelo de forma que se aproxime a una esfera. El término $1 - e^2 (\sin(\phi))^2 \simeq 1$ y dado que $a \gg H$ entonces $a + H \simeq a$. De ahí que la matriz jacobiana pueda verse como

$$\begin{bmatrix} -a \sin(\phi) \cos(\lambda) & -a \cos(\phi) \sin(\lambda) & \cos(\phi) \cos(\lambda) \\ -a \sin(\phi) \sin(\lambda) & a \cos(\phi) \cos(\lambda) & \cos(\phi) \sin(\lambda) \\ a \cos(\phi) & 0 & \sin(\phi) \end{bmatrix} \quad (3.7)$$

La figura 3.3 presenta los valores del jacobiano alrededor de un hemisferio del planeta, es decir desde latitud cero hasta latitud noventa y desde longitud cero hasta longitud de ciento ochenta grados. Se observa, por ejemplo, que la coordenada X y la coordenada Y se ven afectadas por ambos, tanto la latitud como la longitud; la coordenada Z se afecta únicamente por la latitud como era de esperarse dadas las ecuaciones.

Sin embargo, es necesario realizar un análisis local para determinar en el caso de X y Y como se afectarían dada la ubicación en la que estaría operando el sistema. La figura

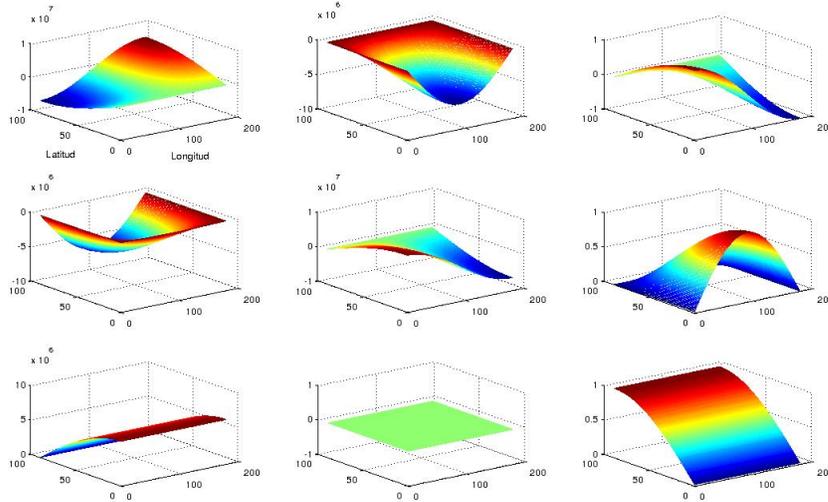


Figura 3.6: Valores del jacobiano alrededor de un hemisferio

3.3 presenta el valor del jacobiano alrededor de una posición local (latitud 21-longitud 89). Un primer análisis muestra que las variaciones en la razón de cambio en una posición local pueden considerarse lineales. En la cercanía de la posición local de la ciudad de Mérida el jacobiano tiene los siguientes valores:

$$\begin{bmatrix} 696.23 & 103,909.87 & 0.0003 \\ 39,887.26 & 1,813.75 & 0.016 \\ 103,925.69 & 0 & 0.0063 \end{bmatrix}$$

Las unidades están dadas en metros por grado de variación. Esto demuestra que para la estimación de la coordenada en X el parámetro mas sensible es la longitud mientras que la latitud afecta significativamente la determinación de la coordenada Y y Z.

En una muestra de 100 mediciones tomadas en una posición fija, el instrumento tiene una incertidumbre determinada por la desviación estándar de las mediciones. La tabla 3.1 muestra la desviación estándar resultante de medir la misma posición con diferentes número de satélites disponibles

3.4. Primera estimación de la homografía

El modelo de cámara en conjunto con la rotación y traslación de una imagen a otra permite encontrar una primera estimación de la homografía entre ambas. Basado en el

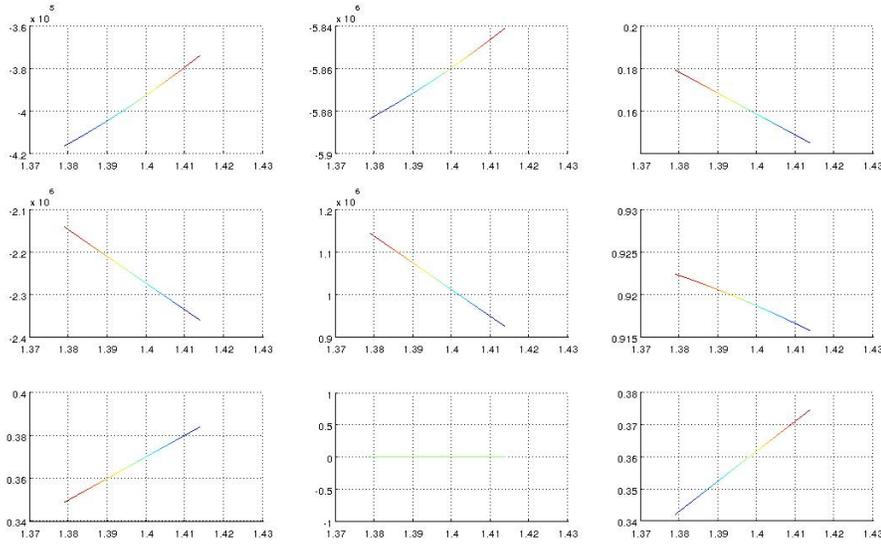


Figura 3.7: Valores del jacobiano alrededor de una posición fija

Cuadro 3.1: Variación en X, Y y Z

| Satélites | X | Y | Z |
|-----------|--------|--------|---------|
| 3 | 4.6955 | 4.7263 | 12.3763 |
| 5 | 3.8648 | 3.8024 | 10.8027 |
| 7 | 1.3772 | 2.6697 | 2.4333 |
| 12 | 0.893 | 1.3443 | 1.245 |

algoritmo DLT (Direct Linear Transformation) que se explicó en el capítulo 2 y, como se mencionó anteriormente, dada la altura de la cámara y que las imágenes corresponden a superficies prácticamente planas (costa yucateca) el mapeo entre imágenes se puede modelar como una homografía.

Suponga que se tiene un punto en el espacio $\mathbf{p} = [X \ Y \ Z \ 1]^T$ en coordenadas homogéneas. De la ecuación 2.4, en un instante t_1 la cámara mapea dicho punto en un plano proyectivo en

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{KDG}_1\mathbf{p} \quad (3.8)$$

Para el mismo punto \mathbf{p} en el espacio, en un instante t_2 , la misma cámara, realiza el

siguiente mapeo

$$\mathbf{x}' = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \mathbf{KDG}_2\mathbf{p} \quad (3.9)$$

Cabe recordar que \mathbf{G} es la matriz en $SE(3)$ que representa la rotación y traslación con respecto al marco global (ecuación 2.3). El algoritmo 3.1 sintetiza el procedimiento para obtener esta homografía.

De esta forma se establecen correspondencias entre pares de imágenes y, como se demostró en el capítulo 2, con cuatro correspondencias se puede aplicar algoritmo DLT para estimar por primera vez la homografía entre ambas imágenes. El refinamiento de dicha homografía se desarrolla en el siguiente capítulo una vez que se han hallado correspondencias directamente entre las imágenes.

Algoritmo 3.1: *Algoritmo para la primera estimación de la homografía.*

1. Establecer cuatro puntos aleatorios en el espacio $p_1 p_2 p_3 p_4$
2. Para cada punto establecer la posición a la que se mapea en la primera posición de la cámara haciendo $\mathbf{x}_n = \begin{bmatrix} x_n \\ y_n \\ 1 \end{bmatrix} = \mathbf{KDG}_1\mathbf{p}_n$ donde \mathbf{K} es la matriz de calibración de la cámara y G_1 es la matriz de rotación y traslación.
3. Para los mismo puntos obtener las correspondencias del mapeo de la cámara en el siguiente instante: $\mathbf{x}'_n = \begin{bmatrix} x'_n \\ y'_n \\ 1 \end{bmatrix} = \mathbf{KDG}_2\mathbf{p}_n$.
4. Aplicar DLT para obtener la primera estimación de la homografía $\hat{\mathbf{H}}$

3.5. Determinación del área de traslape

Una vez obtenida la homografía entre dos imágenes se utiliza esta información para determinar si el área de traslape entre ambas es significativo, esto es importante para la creación del mosaico. En pasos siguientes, será necesario refinar la estimación de la homografía a través de la determinación de correspondencias entre pares de imágenes para lo cual es necesario que exista un área de traslape adecuada. Para determinar esta área se hace uso de la geometría computacional y del cálculo de área de polígonos irregulares de n lados. En el primer caso, de Berg et al. [2008] presenta la deducción para

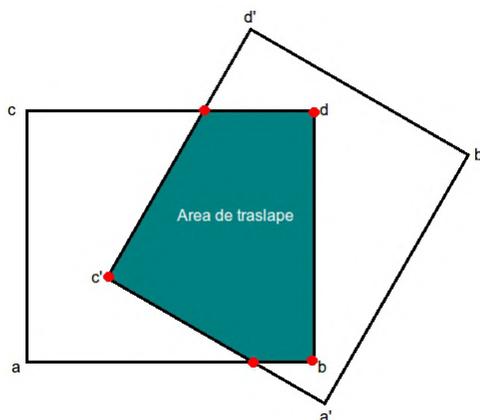


Figura 3.8: Dos imágenes con una área de traslape

obtener las intersecciones en múltiples segmentos de líneas. En el segundo caso, Braden [1986] contiene un excelente resumen de la fórmula de Surveyor necesaria para el cálculo de áreas en polígonos irregulares.

El problema en general se presenta en la figura 3.5, en la cual se puede observar como dos imágenes consecutivas han sido rectificadas para proyectarse en un mismo plano haciendo uso de la homografía determinada anteriormente y se desea conocer el área de traslape entre ambas haciendo uso de los puntos de intersección que conforman un polígono irregular.

3.5.1. Intersección de segmentos

Para la determinación de los puntos de intersección entre dos segmentos de línea se utiliza una técnica de gráficos por computadora en la cual se representan los segmentos por ecuaciones paramétricas. Cualquier punto en un segmento de recta, por ejemplo \overline{ab} de la figura 3.5, puede ser escrito como una combinación convexa usando un parámetro real s de la siguiente forma:

$$p(s) = (1 - s)a + sb$$

de forma similar se introduce un parámetro t para el segmento $\overline{a'c'}$ resultando en

$$q(t) = (1 - t)a' + tc'$$

Una intersección ocurre si y solo si se pueden encontrar valores para s y t en los rangos deseados tales que $p(s) = q(t)$. Por lo que se obtienen dos ecuaciones:

$$\begin{aligned}(1-s)a_x + sb_x &= (1-t)a'_x + tc'_x \\ (1-s)a_y + sb_y &= (1-t)a'_y + tc'_y\end{aligned}$$

Acomodando las ecuaciones se obtiene el sistema de ecuaciones a resolver:

$$\begin{aligned}(s(b_x - a_x) - (c'_x - a'_x)t) &= a'_x - a_x \\ (s(b_y - a_y) - (c'_y - a'_y)t) &= a'_y - a_y\end{aligned}$$

Si la matriz de coeficientes del sistema es singular entonces las líneas son paralelas o colineales.

3.5.2. Área de un polígono irregular

Como se observó en la figura 3.5, el polígono queda determinado no sólo por los puntos de intersección entre los segmentos de línea, sino por los vértices de cada una de las imágenes. Si un vértice se encuentra dentro de la otra imagen se considera vértice del polígono irregular. Una vez determinados vértices es necesario determinar el área del polígono. Para explicar de forma adecuada la manera de llevar a cabo dicho cálculo se presenta el siguiente lema:

Lema 3.5.1 *El valor absoluto del $\det \begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}$ es el área del paralelogramo determinado por los vectores $\mathbf{v} = [v_1 \ v_2]$ y $\mathbf{w} = [w_1 \ w_2]$ y dicho determinante es positivo si la rotación de \mathbf{v} hacia \mathbf{w} es en sentido contrario a las manecillas del reloj.*

dado que un polígono de n lados puede dividirse en $n - 2$ triángulos la fórmula de Surveyor se expresa de la siguiente forma:

Definición 3.5.1 *Si los vértices de un polígono simple, en orden contrario a las manecillas del reloj son $(x_0, y_0), (x_1, y_1), \dots, (x_{n-1}, y_{n-1})$ entonces el área del polígono es:*

$$A = \frac{1}{2} \left\{ \det \begin{bmatrix} x_0 & x_1 \\ y_0 & y_1 \end{bmatrix} + \det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} + \dots + \det \begin{bmatrix} x_{n-2} & x_{n-1} \\ y_{n-2} & y_{n-1} \end{bmatrix} + \det \begin{bmatrix} x_{n-1} & x_0 \\ y_{n-1} & y_0 \end{bmatrix} \right\}$$

El algoritmo 3.2 enumera los pasos necesarios para determinar el área de traslape. La intersección de los bordes siempre formara parte del área de traslape. En el caso de las esquinas de las imágenes se consideran vértices del área de traslape si y solo si las esquinas de una imagen se encuentran dentro de la otra imagen.

Algoritmo 3.2: *Algoritmo para el área de traslape.*

1. Determinar los puntos de intersección de los bordes de las imágenes
2. Determinar si las esquinas de cada imagen son vértices del área de traslape:
3. Ordenar los puntos vértice del polígono en contra de las manecillas del reloj
4. Aplicar la formula de Surveyor para determinar el área

Capítulo 4

Construcción de mosaicos.

El presente capítulo es una revisión teórica de los algoritmos utilizados para llevar a cabo la construcción del mosaico, desde la obtención de los puntos de interés en una imagen, hasta la reproyección de múltiples imágenes en un solo plano proyectivo. La obtención de correspondencias juega un papel muy importante por el hecho de que se aprovecha la primera estimación de la homografía para simplificar y acelerar dicha búsqueda.

La construcción de un mosaico tiene tres partes principales:

- Registro de las imágenes
- Reproyección en un plano.
- Mezclado (*blending*) de las imágenes.

La figura 4 muestra estos pasos.

4.1. Detección de esquinas

Desde su publicación en 1988 el detector de esquinas de Harris (Harris and Stephens [1988]) ha sido ampliamente aceptado y utilizado en muy diversos problemas que involucran la estimación de correspondencias entre dos imágenes. La idea principal del detector de esquinas de Harris es encontrar puntos donde dos bordes se encuentran (gradiente significativo en ambas direcciones), sin embargo, dado que para un punto existe únicamente un valor de gradiente es necesario hacer la valoración sobre una pequeña región o

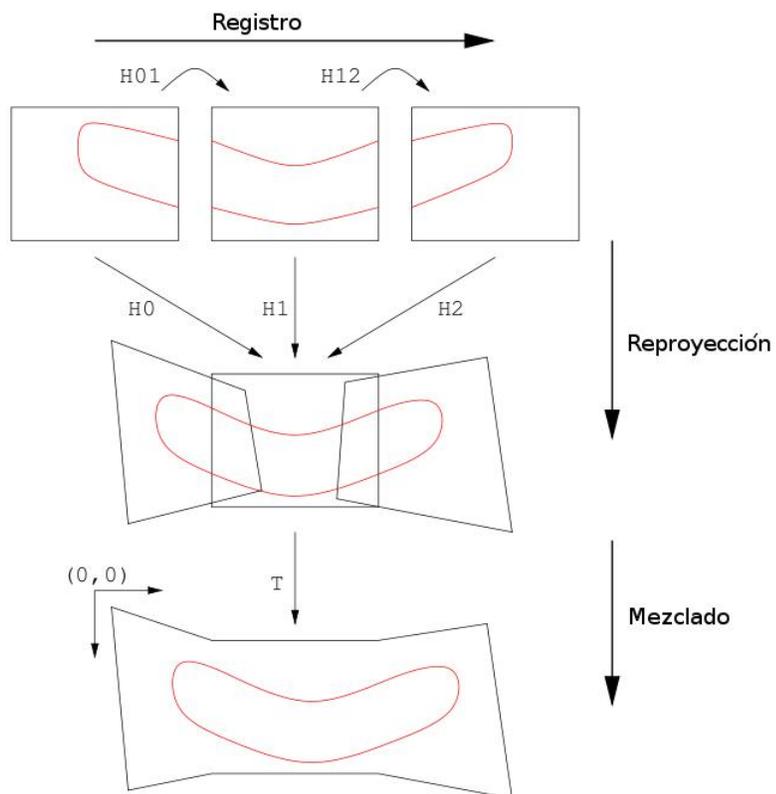


Figura 4.1: Proceso para la formación de un mosaico (Capel [2001])

ventana, si el gradiente es constante existe muy poco o ningún cambio en la intensidad, si el cambio es en una sola dirección se trata entonces de un borde y si existen cambios fuertes en direcciones ortogonales entonces se trata de una esquina.

Sea $I(x, y)$ la función de intensidad de una imagen y $[u, v]$ un corrimiento pequeño, se utiliza la función de autocorrelación que mide los cambios en la imagen en diferentes posiciones:

$$c(x, y) = \sum_i w(x_i, y_i) [I(x_i + u, y_i + v) - I(x_i, y_i)]^2 \quad (4.1)$$

donde $w(x_i, y_i)$ es una función de ventana que generalmente es una Guassiana.

La imagen con corrimiento se puede aproximar a través de la expansión de Taylor truncada en los términos de primer orden.

$$I(x_i + u, y_i + v) \approx I(x_i, y_i) + uI_x(x_i, y_i)vI_y(x_i, y_i) \quad (4.2)$$

donde I_x e I_y representan las derivadas con respecto a x y y respectivamente.

Sustituyendo 4.1 en 4.2 se tiene que

$$\begin{aligned} c(x, y) &\approx \sum_i w(x_i, y_i) [I(x_i, y_i) + uI_x(x_i, y_i)vI_y(x_i, y_i) - I(x_i, y_i)]^2 \\ &= \sum_i w(x_i, y_i) [u^2I_x^2 + 2uvI_xI_y + v^2I_y^2] \\ &= \sum_i w(x_i, y_i) \left([u, v] \begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \right) \\ &= \left([u, v] \sum_i w(x_i, y_i) \begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \right) \end{aligned} \quad (4.3)$$

Las esquinas o puntos de interés se encuentran precisamente con los eigenvalores (λ_1, λ_2) de la matriz

$$\mathbf{Q}(x, y) = \sum_i w(x_i, y_i) \begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix} = \sum_i w(x_i, y_i) \begin{bmatrix} A & B \\ B & C \end{bmatrix}$$

de acuerdo a lo siguiente:

- \mathbf{Q} es positiva definida si $\lambda_1, \lambda_2 > 0$.

- $\lambda_1\lambda_2 = \det \mathbf{Q} = AC - B^2$, $\lambda_1 + \lambda_2 = \text{traza } \mathbf{Q} = A + C$.
- De acuerdo con Harris and Stephens [1988] el valor de un pixel como esquina esta dado por $E = \lambda_1\lambda_2 - (\lambda_1 + \lambda_2)^2$. Luego entonces, una esquina es un valor máximo de E .

El algoritmo 4.1 resume los pasos necesarios para implementar este detector de esquinas.

Algoritmo 4.1: *Algoritmo de detección de esquinas de Harris.*

1. Suavizar la imagen para reducir el ruido aplicando un filtro Gaussiano con $\sigma \sim 1$.
2. Calcular las derivadas I_x , I_y .
3. Para cada punto en la imagen (x, y) y sobre una ventana W :
 - i $A = \sum_W I_x^2$, $B = \sum_W I_x I_y$, $C = I_y^2$.
 - ii Obtener el valor de esquina $E = \lambda_1\lambda_2 - (\lambda_1 + \lambda_2)^2$.
4. Suprimir los valores no máximos en una pequeña vecindad de acuerdo a un valor τ para obtener los pixeles que son esquinas.

4.2. Correspondencias

Una vez obtenidos los puntos de interés es necesario determinar las correspondencia de estos puntos entre un par de imágenes. Uno de los procedimientos mas comunes (?) es el de *correlación cruzada* que consiste en hallar un coeficiente de correlación entre dos vecindades, la imagen de referencia (A) y la imagen a verificar (B).

$$c = \frac{\sum_{i=0}^m \sum_{j=0}^n [(A_{ij} - \bar{A})(B_{ij} - \bar{B})]}{\sqrt{\left[\sum_{i=0}^m \sum_{j=0}^n (A_{ij} - \bar{A})^2 \right] \left[\sum_{i=0}^m \sum_{j=0}^n (B_{ij} - \bar{B})^2 \right]}} \quad (4.4)$$

Si no se cuenta con información acerca de la transformación entre ambas imágenes la búsqueda debe ser una comparación de cada punto de interés de la imagen A con todos los puntos de interés de la imagen B siendo el punto que mayor correlación tenga considerado como una correspondencia. Una forma de reducir dicha búsqueda consiste en suponer que la transformación entre las imágenes es pequeña por lo que se compararían únicamente los puntos de interés en un área pequeña centrada en dicho punto.

Sin embargo, si se conoce la relación de transformación que mapea los píxeles de \mathbf{A} con los de \mathbf{B} se puede realizar una búsqueda dedicada sobre un ventana pequeña. El algoritmo 4.2 presenta la forma en que se lleva a cabo dicho proceso.

Algoritmo 4.2: *Algoritmo de obtención de correspondencias*

1. Sea \mathbf{A} y \mathbf{B} imágenes con puntos de interés identificados y de las que se ha estimado la matriz de homografía $\hat{\mathbf{H}}$ con anterioridad.
2. Para cada punto de interés $\mathbf{p}_k = (x_k, y_k)$ en \mathbf{A} hacer:
 - i $\mathbf{p}'_k = \mathbf{p}_k \hat{\mathbf{H}}$.
 - ii Construir una ventana W al rededor de \mathbf{p}'_k en \mathbf{B}
 - iii Para cada punto en W calcular la correlación para una región de 3×3

$$c = \frac{\sum_{i=0}^m \sum_{j=0}^n [(A_{ij} - \bar{A})(B_{ij} - \bar{B})]}{\sqrt{\left[\sum_{i=0}^m \sum_{j=0}^n (A_{ij} - \bar{A})^2 \right] \left[\sum_{i=0}^m \sum_{j=0}^n (B_{ij} - \bar{B})^2 \right]}}$$

3. Si la correlación de mayor valor en W , $c_{max} > 9.0$ dicho píxel se considera correspondencia de \mathbf{p}_k

4.3. Segunda estimación de la homografía

Para la primera estimación de la homografía entre dos imágenes se utilizaron únicamente 4 correspondencias obtenidas del modelo de cámara y de la orientación entre un par de imágenes. Sin embargo, esta primera estimación es utilizada para optimizar el proceso de obtención de correspondencias entre las imágenes y es necesario realizar una estimación mas robusta de la homografía existente. Para ello se implementa un algoritmo denominado RANSAC (*Random Sample Consensus*) que tiene la propiedad de eliminar las correspondencias falsas que en procedimiento anterior pudieran determinarse. Una referencia clásica en este método se encuentra en Hartley and Zisserman [2004].

Consideremos entonces una serie de puntos \mathbf{x}_i en una imagen y sus correspondencias \mathbf{x}'_i en la otra. Se desea estimar la homografía \mathbf{H} tal que $\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i$ donde \mathbf{H} es de la forma:

$$\mathbf{H} = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix} \quad (4.5)$$

Este modelo requiere de cuatro correspondencias para su solución, sin embargo, se

tiene un número $n \gg 4$ de correspondencias de las cuales hay un porcentaje que son incorrectas y no se ajustan al modelo real de homografía entre las imágenes. Si se toman todas las mediciones y se estime la homografía se tendrá un error que dependerá en gran medida de la cantidad y dimensión del ruido de la medición, además de los errores que puede tener el método escogido para dicha estimación. Es por ello, que el primer paso es eliminar la mayor cantidad posible de mediciones erróneas para proceder a estimar la homografía. RANSAC realiza este proceso de selección de los mejores candidatos a formar parte de dicha estimación, rechazando completamente los que no son aptos para ello. La forma de hacerlo es tomar de forma aleatoria el menor número de correspondencias necesarias para estimar la homografía ($n = 4$) y resolver \mathbf{H} . Una vez hecho esto, se mide una distancia al modelo de cada una de las demás correspondencias y se consideran adecuadas las que están a una distancia menor a t e incorrectos los que se encuentran mas lejos; se debe iterar una cantidad de veces y después escoger como primera estimación de \mathbf{H} la que haya tenido la mayor cantidad de correspondencias adecuadas para utilizarlas en mejorar dicha estimación, generalmente, y en este caso, minimizando una función de costo (mínimos cuadrados). Se puede agregar un paso más de incrementar el número de correspondencias con la homografía estimada para re-calcularla mejorando así la estimación.

Como puede observarse existen tres partes importantes en el algoritmo:

1. Estimar la homografía con tan solo 4 correspondencias aleatorias.
2. Determinar y calcular las distancias a dicha estimación.
3. Usar un método numérico para minimizar una función de mínimos cuadrados.

El primer paso se describió para la primera estimación de la homografía y consiste en la aplicación del algoritmo DLT.

No es necesario realizar un gran número de iteraciones para encontrar un conjunto numeroso de correspondencias que se ajustan a un modelo determinado por las cuatro correspondencias aleatorias. Existe un método adaptativo para detener la búsqueda de dicho conjunto. En vez de eso un número N de muestras son tomadas asegurando con una probabilidad p de que, al menos, una de ellas está libre de correspondencias

erróneas. Supongamos que w es la probabilidad de que la muestra tenga únicamente correspondencias correctas por lo tanto $\epsilon = 1 - w$ es la probabilidad de que sea incorrecta. Entonces, al menos N muestras de cuatro correspondencias cada una son necesarias, donde $(1 - w^4)^N = 1 - p$, así que

$$N = \frac{\log(1 - p)}{\log(1 - (1 - \epsilon)^4)}$$

En este caso se utilizó $p = 0.99$ y $\epsilon = 1 - (\text{inliers/correspondencias})$.

La distancia se midió a través de la transferencia simétrica $d(x_i, H^{-1}x'_i)^2 + d(x'_i, Hx_i)^2$. Se utilizó un umbral de $t = 15$ para determinar las correspondencias adecuadas.

El tercer paso requiere de plantear el problema de minimización al que se enfrenta. Teniendo un conjunto de correspondencias se desea hallar la homografía minimizando una función de costo. En este caso se utiliza una función de distancia geométrica llamada error de transferencia simétrica definido como $\sum_i d(x_i, H^{-1}x'_i)^2 + d(x'_i, Hx_i)^2$ donde $d(\cdot)$ representa la distancia euclidiana. Esta función puede representarse como un problema no lineal de mínimos cuadrados:

$$C(\mathbf{p}) = \| \mathbf{X} - F(\mathbf{p}) \|^2 \quad (4.6)$$

Aquí \mathbf{X} puede ser los cuatro puntos de una correspondencia y $F(\mathbf{p})$ una función no lineal con parámetros \mathbf{p} como la de transferencia simétrica. Una vez que logremos definir concretamente los elementos de la ecuación 4.6 existen varios métodos iterativos para hallar la solución.

Para todas las correspondencias (x_i, x'_i) donde $i = 1, 2, \dots, n$ y n es el número de correspondencias hacemos

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad X' = \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix} \quad H^{-1}X' = \begin{bmatrix} H^{-1}x'_1 \\ H^{-1}x'_2 \\ \vdots \\ H^{-1}x'_n \end{bmatrix} \quad HX = \begin{bmatrix} Hx_1 \\ Hx_2 \\ \vdots \\ Hx_n \end{bmatrix}.$$

Cabe recordar que $H = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}$ y se denotará $H^{-1} = \begin{bmatrix} h'_1 & h'_2 & h'_3 \\ h'_4 & h'_5 & h'_6 \\ h'_7 & h'_8 & h'_9 \end{bmatrix}$

por simplicidad de notación, aunque cada elemento de la matriz invertida contiene únicamente valores de la matriz H original. Todo esto nos conduce a que :

$$\begin{aligned}
C(H) &= \sum_i d(x_i, H^{-1}x'_i)^2 + d(x'_i, Hx_i)^2 \\
&= \| X - H^{-1}X' \|^2 + \| X' - HX \|^2 \\
&= \left(x_i - \frac{h'_1x'_i + h'_2y'_i + h'_3}{h'_7x'_i + h'_8y'_i + h'_9} \right)^2 + \left(y_i - \frac{h'_4x'_i + h'_5y'_i + h'_6}{h'_7x'_i + h'_8y'_i + h'_9} \right)^2 \\
&\quad + \left(x'_i - \frac{h_1x_i + h_2y_i + h_3}{h_7x_i + h_8y_i + h_9} \right)^2 + \left(y'_i - \frac{h_4x_i + h_5y_i + h_6}{h_7x_i + h_8y_i + h_9} \right)^2
\end{aligned} \tag{4.7}$$

Si hacemos $\mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_9 \end{bmatrix}$ la función de costo a minimizar será $C(H) = C(\mathbf{h}) = f(\mathbf{h})^T f(\mathbf{h})$ donde

$$f(\mathbf{h}) = \begin{bmatrix} x_1 - \frac{h'_1 x'_1 + h'_2 y'_1 + h'_3}{h'_7 x'_1 + h'_8 y'_1 + h'_9} \\ y_1 - \frac{h'_4 x'_1 + h'_5 y'_1 + h'_6}{h'_7 x'_1 + h'_8 y'_1 + h'_9} \\ x'_1 - \frac{h_1 x_1 + h_2 y_1 + h_3}{h_7 x_1 + h_8 y_1 + h_9} \\ y'_1 - \frac{h_4 x_1 + h_5 y_1 + h_6}{h_7 x_1 + h_8 y_1 + h_9} \\ x_2 - \frac{h'_1 x'_2 + h'_2 y'_2 + h'_3}{h'_7 x'_2 + h'_8 y'_2 + h'_9} \\ y_2 - \frac{h'_4 x'_2 + h'_5 y'_2 + h'_6}{h'_7 x'_2 + h'_8 y'_2 + h'_9} \\ x'_2 - \frac{h_1 x_2 + h_2 y_2 + h_3}{h_7 x_2 + h_8 y_2 + h_9} \\ y'_2 - \frac{h_4 x_2 + h_5 y_2 + h_6}{h_7 x_2 + h_8 y_2 + h_9} \\ \vdots \\ x_n - \frac{h'_1 x'_n + h'_2 y'_n + h'_3}{h'_7 x'_n + h'_8 y'_n + h'_9} \\ y_n - \frac{h'_4 x'_n + h'_5 y'_n + h'_6}{h'_7 x'_n + h'_8 y'_n + h'_9} \\ x'_n - \frac{h_1 x_n + h_2 y_n + h_3}{h_7 x_n + h_8 y_n + h_9} \\ y'_n - \frac{h_4 x_n + h_5 y_n + h_6}{h_7 x_n + h_8 y_n + h_9} \end{bmatrix} \quad (4.8)$$

El gradiente de la función de costo esta determinado por $\nabla C(\mathbf{h}) = J_f(\mathbf{h})^T f(\mathbf{h})$ donde $J_f(\mathbf{h})$ es el jacobiano de $f(\mathbf{h})$, una matriz de $4n \times 9$. El Hessiano esta determinado por $C''(\mathbf{h}) = J_f(\mathbf{h})^T J_f(\mathbf{h})$. Con estos datos es posible implementar varios métodos de optimización.

El algoritmo de Levenberg-Maquardt (LM) es un híbrido del descenso de gradiente y el Gauss-Newton y la idea es tomar lo mejor de cada uno de ellos en el transcurso del algoritmo. Cuando se está lejos de la solución prácticamente el algoritmo funciona como un descenso de gradiente bajando lentamente hacia el mínimo, sin embargo, al estar cerca de la solución se dan saltos grandes como en el método de Gauss-Newton.

En cada iteración se calcula de nuevo los parámetros de \mathbf{h} como sigue

$$\mathbf{h}_{k+1} = \mathbf{h}_k - \mathbf{h}_k - (J_f(\mathbf{h})^T J_f(\mathbf{h}) + \mu_k I)^{-1} J_f(\mathbf{h})^T f(\mathbf{h}) \quad (4.9)$$

donde I es la matriz identidad y μ_k es un escalar de manera que se asegure que la estimación del Hessiano sea semipositiva definida.

Para la primera iteración se puede determinar μ_0 de la siguiente forma

$$\mu_0 = \tau \times \max(\text{diag } He) \quad (4.10)$$

donde $0 < \tau \ll 1$. Después en cada iteración el valor de μ se actualiza. Primero se calcula la razón de reducción de la función de costo

$$\rho = \frac{C(\mathbf{h}_k) - C(\mathbf{h}_{k+1})}{\delta_h^T (\mu_k \delta_h - J_f(\mathbf{h}_k)^T f(\mathbf{h}_k))} \quad (4.11)$$

donde $\delta_h = \mathbf{h}_{k+1} - \mathbf{h}_k$.

Si $\rho > 0$

$$\mu_{k+1} = \mu_k \times \max\left(\frac{1}{3}, 1 - (2\rho - 1)^3\right) \quad (4.12)$$

Si $\rho \leq 0$ en esa iteración de hecho se ha incrementado la función de costo en vez de minimizarla por lo que $\mathbf{h}_{k+1} = \mathbf{h}_k$ y

$$\mu_{k+1} = 2\mu_k \quad (4.13)$$

La figura 4.2 muestra el resultado de aplicar el algoritmo en dos imágenes en donde la cámara ha sufrido una rotación sobre su propio eje. La imagen izquierda muestra los puntos de interés que se encontraron y la de la derecha las correspondencias una vez aplicado RANSAC. En la parte de abajo se presenta la imagen corregida geoméricamente utilizando la homografía estimada con anterioridad.

4.4. Registro geométrico

El objetivo del registro geométrico es que todas las imágenes se encuentren en un solo marco de referencia global. La elección de dicho marco es indiferente y generalmente

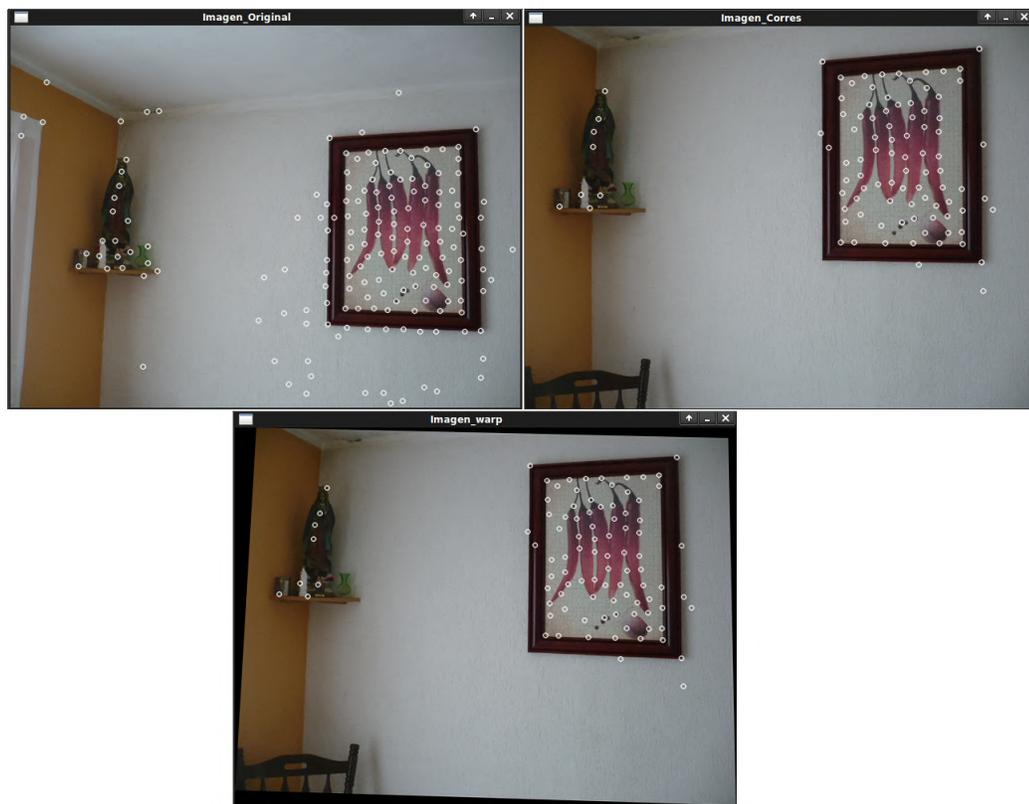


Figura 4.2: Ejemplo de estimación de homografías con RANSAC



Figura 4.3: Proyección en una superficie plana

se toma una imagen centrada en la escena capturada. Se considera que la *imagen de referencia* esta alineada con los ejes del marco global por lo que la transformación entre el marco global y la imagen de referencia es la identidad (Capel [2001]). Una vez que se han obtenido las homografías entre cada imagen consecutiva, la transformación entre una imagen en particular y la imagen de referencia es obtenida mediante la concatenación de las homografías intermedias de la secuencia. Sea m_k una imagen en particular y m_{ref} la imagen de referencia la homografía que transforma la primera en el marco de la segunda está dada por

$$\mathbf{H}_{(ref \rightarrow k)} = \prod_{n=ref}^{k-1} \mathbf{H}_{(n \rightarrow n+1)}$$

4.5. Proyección en un plano

La forma mas sencilla de reproyectar las imágenes es un plano. Esto es, la función de mapeo es la misma homografía. Esta técnica es funcional para condiciones donde la cámara ha sufrido una rotación sobre el eje óptico o bien en escenas planas como es el caso de las imágenes aéreas. Otra ventaja de esta proyección es que no se requiere de mayor información que la homografía entre imágenes a diferencia de otras proyecciones que requieren de la matriz de calibración. El resultado de esta proyección es una imagen con una distorsión proyectiva en los bordes de la imagen donde las líneas se preservan como si se tratará de un sensor extremadamente amplio. En la figura 4.5 se aprecia la proyección sobre un plano de una secuencia de imágenes y en la figura 4.5 sobre un cilindro.



Figura 4.4: Proyección en una superficie cilíndrica

4.6. Mezclado

De acuerdo con Capel [2001], cada pixel debe ser tratado de la siguiente forma:

1. Usando la homografía encontrada y de acuerdo con el plano de reproyección elegido transformar cada pixel de cada imagen.
2. Para cada pixel de cada imagen transformada que se encuentre dentro del área de reproyección del mosaico calcular su intensidad usando una técnica adecuada de interpolación. Para este caso, interpolación bilineal.
3. Combinar el conjunto de intensidades para un pixel del mosaico en un único valor usando alguna función de mezclado.

Para el mezclado existen diferentes funciones que van desde un simple promedio de intensidades hasta filtros. En este trabajo se utilizó la función de desvanecimiento de bordes. Es una función que otorga pesos a los pixeles siendo los de mayor peso los que se encuentran cercanos al centro de la imagen y los de menor los cercanos a los bordes. De esta forma, al momento de calcular los promedios de pixeles se consideran dichos pesos. La función bicuadrática usada para esta fin es:

$$f(x, y) = \left(1 - \left(x - \frac{w}{2}\right)^2\right) \left(1 - \left(y - \frac{h}{2}\right)^2\right)$$

4.7. Resumen

En resumen, los pasos necesarios para la construcción del mosaico son como sigue:

- Determinar la matriz en $SE(3)$ que transforma cada posición de la cámara a un marco global, en este caso, la primera posición de la persona en tierra, donde dicho marco tenga el eje Z ortogonal a la superficie terrestre y el eje Y orientado hacia el norte.
- Para cada imagen determinar la homografía con respecto a la anterior haciendo uso del procedimiento descrito en el capítulo 3.
- Una vez obtenida la homografía, determinar el área de traslape. En caso de que sea menor al 40% o mayor al 60% de la imagen anterior se debe descartar adquirir una nueva imagen en ese instante.
- Ya con las imágenes obtenidas, realizar el refinamiento de la homografía entre pares de imágenes consecutivas. Para ello determinar las correspondencias usando la primera homografía y aplicar RANSAC.
- Realizar la corrección geométrica usando la homografía usando la primera imagen como referencia
- Para cada imagen determinar el peso de cada pixel de acuerdo a su cercanía con el borde
- En la imagen del mosaico realizar un promedio ponderado con los pesos anteriores para calcular el valor del pixel en el mosaico.

Capítulo 5

Trabajo experimental.

5.1. Primera estimación de la homografía

El primer paso en el trabajo experimental consistió en obtener la matriz de calibración \mathbf{K} de la cámara, en este caso una cámara digital marca Benq modelo DC C35 de 3.2 Megapíxeles. Se utilizaron 12 fotografías para la calibración, la figura 5.1 muestra dichas imágenes.

Usando el *toolbox* de Matlab creado por Jean-Yves Bouguet para la calibración de cámaras se obtuvo la siguiente matriz de calibración que proporciona información de la distancia focal y las coordenadas del punto principal de la cámara.

$$\mathbf{K} = \begin{bmatrix} 2582.49 & 0 & 1023.5 \\ 0 & 2574.058 & 767.5 \\ 0 & 0 & 1 \end{bmatrix}$$

El segundo paso fue la adquisición de las imágenes aéreas para lo cual se levantó el sistema a una altura de 125 metros en el norte de la ciudad de Mérida y se obtuvieron imágenes y datos de posición global. Se obtuvieron cuatro imágenes adecuadas para usar en el presente trabajo (las demás estaban fuera de foco debido al movimiento de la cámara) como se observa en la figura 5.2. La tabla 5.1 muestra a su vez las posiciones globales de la cámara para cada una de las fotos, donde *aire1* es la imagen de la esquina superior izquierda y la numeración continua en contra de las manecillas del reloj hasta la imagen *aire4*. La última fila muestra las coordenadas del punto en tierra donde fueron tomadas todas la imágenes.

Se tomaron dos imágenes, *aire2* y *aire3* para demostrara experimentalmente el planteamiento.

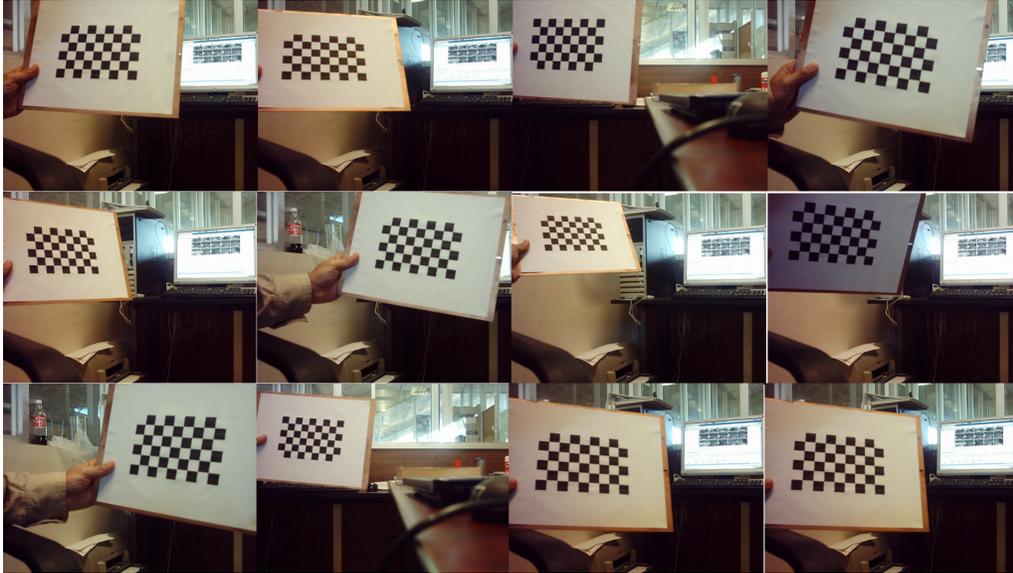


Figura 5.1: Imágenes utilizadas para la calibración de la cámara

Cuadro 5.1: Coordenadas de posición de las cuatro imágenes

| Imagen | Latitud | Longitud | Altitud | Geopeso |
|---------|-----------|-----------|---------|---------|
| aire1 | 21.035575 | 89.633728 | 112 | -10.5 |
| aire2 | 21.035353 | 89.633756 | 115 | -10.5 |
| aire3 | 21.035103 | 89.633594 | 118 | -10.5 |
| aire4 | 21.035823 | 89.633701 | 102 | -10.5 |
| control | 21.037006 | 89.632367 | 6 | -10.5 |



Figura 5.2: Secuencia de cuatro imágenes capturadas con su posición global al norte de la ciudad de Mérida

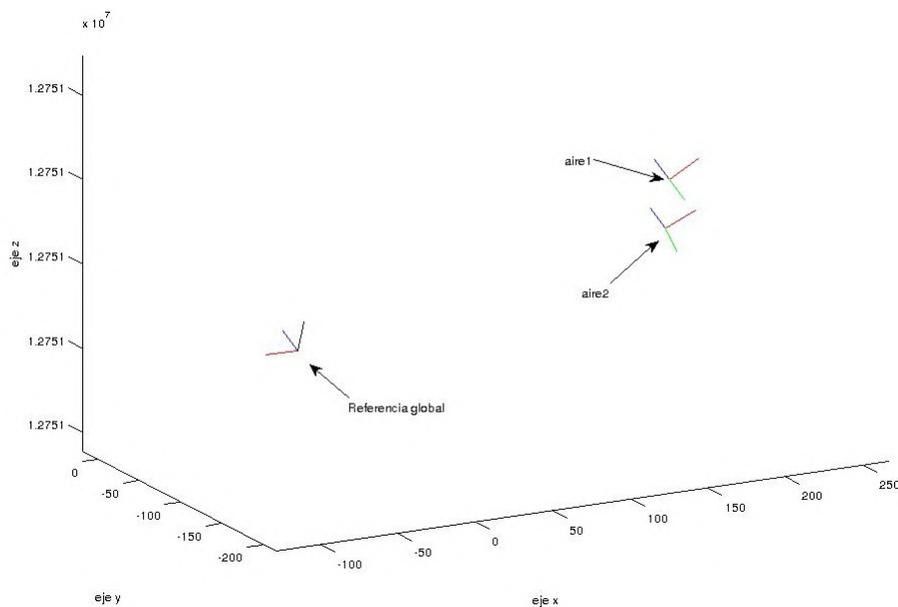


Figura 5.3: Los marcos locales de la cámara al momento de tomar la imagen para *aire2* y *aire3*.

Se utilizó el mismo procedimiento para corroborar que las transformaciones entre marcos de referencia y marcos locales de la cámara dieran como resultado la rotación sobre el eje Z que existiría entre imágenes. La figura 5.3 muestra la posición relativa entre el marco de la imagen y el marco global utilizado. Dichos marcos ya han sido transformados para que el eje Z sea perpendicular a la superficie terrestre y el eje Y se encuentre orientado precisamente hacia la posición de la persona en tierra, correspondiente a la línea del papalote. Esta misma transformación es presentada en la imagen 5.4 pero con las imágenes obtenidas y puede observarse como el ángulo de captura se vuelve paralelo a la línea de t

Ya en la figura 5.3 se puede apreciar la rotación que se tendría que llevar a cabo para sin embargo, para dar mayor claridad a la idea, la figura 5.5 muestra la forma en que están orientados ambos marcos locales si no existiese traslación entre ellos.

A manera de demostración se toman dos imágenes de la secuencia, *aire2* y *aire3*, como se muestran en la figura 5.6 en donde se observa la rotación que sufrió la cámara entre imagen e imagen.

Después de aplicar las operaciones descritas en el capítulo 3 se encuentra la matriz

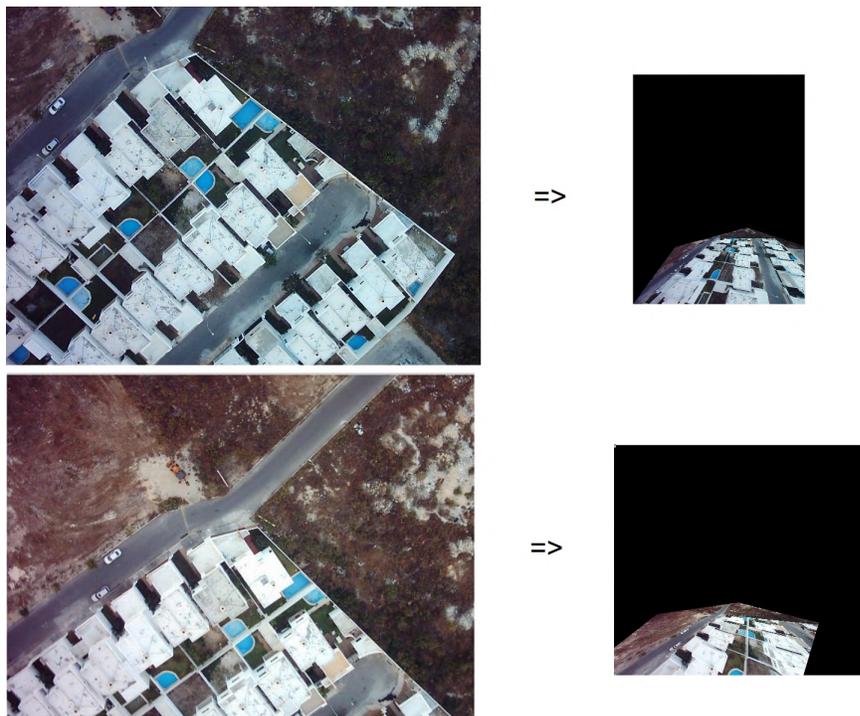


Figura 5.4: Las imágenes *aire2* y *aire3* transformadas al marco global usado como referencia

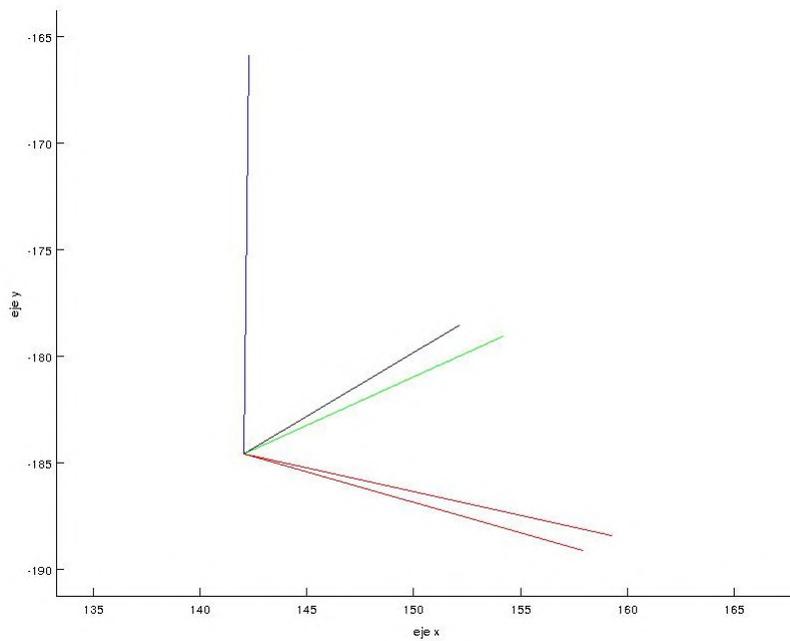


Figura 5.5: Rotación de la cámara si los marcos locales de *aire2* y *aire3* estuviesen en la misma posición



Figura 5.6: Dos imágenes con área de traslape suficiente y rotación de la cámara entre ambas

Cuadro 5.2: Correspondencias para cuatro puntos del espacio y dos instantes de la cámara

| \mathbf{X} | \mathbf{x} | \mathbf{x}' |
|------------------------|-------------------------------|-------------------------------|
| $[23 \ 2332 \ 23]^T$ | $[82937.96 \ 6020356.14]^T$ | $[-596705.26 \ 6067792.24]^T$ |
| $[1233 \ 43 \ 3]^T$ | $[3187291.25 \ 11298.70]^T$ | $[3213991.42 \ 590333.12]^T$ |
| $[2 \ 10 \ 54]^T$ | $[60433.99 \ 67185.58]^T$ | $[122014.87 \ 153728.91]^T$ |
| $[1533 \ 2233 \ 31]^T$ | $[3990698.82 \ 5771664.38]^T$ | $[3312768.68 \ 6301227.02]^T$ |

de transformación en $SE(3)$ que tuvo la cámara en el sentido de *aire3* hacia *aire2*.

$$\mathbf{G}_1 = \begin{bmatrix} 0.992339 & -0.123538 & 0 & 17.19419 \\ 0.123538 & 0.992339 & 0 & 27.50592 \\ 0 & 0 & 1 & 19.938313 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Se toma como marco de referencia local el de la cámara al momento de adquirir la imagen *aire2*. Por lo tanto, la matriz en dicho marco es la identidad.

$$\mathbf{G}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Se eligen cuatro puntos en el espacio tridimensional para aplicar las ecuaciones 3.8 y 3.9 y así obtener las correspondencias entre imágenes. La única condición es utilizar cuatro puntos que no sean colineales y que se encuentren a suficiente distancia formando preferentemente un rectángulo. La tabla 5.2 presenta el resultado de este proceso.

Usando el algoritmo 2.1 se obtiene la matriz de homografía entre ambas imágenes:

$$\mathbf{H} = \begin{bmatrix} 0.196590 & -0.025242 & -410.398071 \\ 0.020250 & 0.199122 & -813.902161 \\ -0.000002 & -0.000004 & 1 \end{bmatrix}$$

La figura 5.7 muestra la imagen *aire3* rectificadas al plano de la imagen *aire2*. La figura 5.8 es una superposición de las imágenes a modo de mosaico. El resultado es una estimación adecuada de la matriz de transformación entre las imágenes.

5.2. Conclusiones

En el presente trabajo se presenta una metodología para la estimación de la homografía entre dos imágenes a partir de la posición global de la cámara en ambos instantes.



Figura 5.7: Aplicación de la matriz de homografía encontrada (prácticamente una rotación) a una de las imágenes de la secuencia



Figura 5.8: Superposición de dos imágenes (referencia y transformada) a manera de mosaico

Junto con la posición global de la persona en tierra que controla el sistema de adquisición es posible plantear un problema de tipo geométrico y algebraico con el cual, a través de matrices de rotación, se obtiene la matriz de movimiento de cuerpo rígido en $SE(3)$. Una vez obtenida dicha matriz, se utiliza para obtener correspondencias entre ambas imágenes haciendo uso del modelo de cámara *pinhole* para lo cual fue necesario realizar la calibración de la cámara. La matriz de calibración y la matriz de transformación en $SE(3)$ constituyen los parámetros intrínsecos y extrínsecos respectivamente de la cámara lo que permite saber la posición en que se mapea cualquier punto en el espacio. con cuatro puntos aleatorios se pueden obtener correspondencias entre las imágenes y con ello estimar, a través del algoritmo DLT, la homografía entre ambas.

Además, se presenta la metodología para la estimación del área de traslape entre ambas imágenes. Una vez obtenida la homografía se transforman las esquinas de la segunda imagen para establecer su posición en el plano de la imagen de referencia estableciendo un problema puramente geométrico. Al hallar los puntos de intersección de los lados de cada imagen se forma un polígono irregular de n lados que dependerán de las intersecciones cuya área puede obtenerse a través de la fórmula de Surveyor.

En el trabajo experimental se tomaron imágenes en el norte de la ciudad de Mérida a una altura de 135 metros. Aun cuando el sistema de adquisición presentó problemas de diseño fue posible recuperar la posición global de cada imagen, lo que permitió llevar a cabo la demostración del modelo propuesto para obtener la rotación sobre el eje Z entre las imágenes. La información del acelerómetro se perdió por lo que no se pudo realizar la rectificación sobre los ejes X y Y , sin embargo, se escogieron cuatro imágenes de la secuencia que presentaran poca o nula rotación en dichos ejes.

En cuanto al sistema de adquisición, se utilizó un papalote comercial tipo *Rokkaku* que proporciona estabilidad y facilidad de vuelo al aire libre. Sin embargo, dado que el modelo es antiguo, la cámara sufrió con los movimientos laterales debido al viento y los movimientos verticales al momento de soltar más cuerda. un gran número de imágenes estaban fuera de foco por lo que no pudieron usarse.

El sistema de adquisición se reiniciaba constantemente por lo que se perdían los datos anteriores. El reinicio se debió a ruido eléctrico que el relevador de disparo implementado introducía al circuito del Arduino. Dado que la información recabada del GPS y del

acelerometro era almacenada en memoria RAM, lo datos se perdían en cada reinicio. A la altura en la que se encontraba el equipo y dado que estaba en espacio abierto se lograron obtener hasta 15 satélites de referencia para el calculo de la posición por lo que la estabilidad

Es necesario, como trabajo futuro, construir un sistema de adquisición más robusto con mayor detalle y un gabinete donde las tarjetas se encuentren sujetadas firmemente para evitar desconexiones como ocurrió en un par de ocasiones. Incluir en el programa de la tarjeta la obtención en tiempo real de la homografía así como el área de traslape e integrar la información del acelerometro como parte de los parámetros para determinar si es conveniente adquirir una imagen es un determinado instante ya que con un poco de movimiento la cámara se salía de foco.

Como se muestra en el apéndice puede usarse el filtro de Kalman para integrar la información de los sensores del sistema de modo que se consideren en el modelo las varianzas de cada uno en la estimación de la homografía.

Apéndice A

Filtro de Kalman

Para mejorar la estimación de la posición de la cámara, sobre lo cual se basa todo el trabajo realizado, se implementó un filtro de Kalman. Orr [1992] y Welch and Bishop [1995] presentan un guía muy simple y a la vez ilustrativa de las bases del filtro y de las ecuaciones para implementarlo sobre funciones no lineales a través del Filtro Extendido de Kalman.

Es un algoritmo de procesamiento de datos óptimo recursivo. Óptimo porque minimiza un criterio determinado y porque incorpora toda la información que se le suministra para determinar el filtrado. Recursivo porque no precisa mantener los datos previos, lo que facilita su implementación en sistemas de procesamiento en tiempo real. Por último, algoritmo de procesamiento de datos, ya que es un filtro, pensado para sistemas discretos. El objetivo del filtro de Kalman es estimar los estados de una manera óptima, de manera que se minimiza el índice del error cuadrático medio.

Si se considera \mathbf{a} como un estado verdadero y \mathbf{x} como una observación perfecta en entonces la ecuación que describe matemáticamente como se relacionan las observaciones y los estados es:

$$\mathbf{x} = \mathbf{M}\mathbf{a}$$

Dicha ecuación es conocida como la ecuación de medida o modelo del sensor en una forma lineal. El filtro de Kalman procesa cada observación imperfecta con media $\hat{\mathbf{x}}$, covarianza \mathbf{A} y ecuación de medida para dar como resultado una estimación nueva con media $\hat{\mathbf{a}}$ y covarianza \mathbf{S} .

Cada nueva estimación es una actualización de la anterior corregida por una cantidad proporcional al error actual $(\hat{\mathbf{x}}_i - \mathbf{M}_i\hat{\mathbf{a}}_{i-1})$ llamado la *innovación*. El factor de proporcionalidad \mathbf{K}_i es llamado la *ganancia de Kalman* resultando la estimación en

$$\hat{\mathbf{a}}_i = \hat{\mathbf{a}}_{i-1} + \mathbf{K}_i(\hat{\mathbf{x}}_i - \mathbf{M}_i\hat{\mathbf{a}}_{i-1}) \quad (\text{A.1})$$

$$\mathbf{S}_i = (\mathbf{I} - \mathbf{K}_i\mathbf{M}_i)\mathbf{S}_{i-1} \quad (\text{A.2})$$

donde $\mathbf{K}_i = \mathbf{S}_{i-1}\mathbf{M}_i^T(\mathbf{A}_i + \mathbf{M}_i\mathbf{S}_{i-1}\mathbf{M}_i^T)^{-1}$

En el caso de que la función de medida sea no lineal (la mayoría de los casos son así) se puede linealizar a través de la serie de Taylor truncada en el término de primer grado. Por ejemplo para una función no lineal

$$\mathbf{f}_i(\mathbf{x}_i, \mathbf{a}) = 0$$

se consideran las medias $\hat{\mathbf{x}}_i$ y $\hat{\mathbf{a}}_{i-1}$ como buenas aproximaciones entonces

$$\mathbf{f}_i(\mathbf{x}_i, \mathbf{a}) = 0 \approx \mathbf{f}_i(\hat{\mathbf{x}}_i, \hat{\mathbf{a}}_{i-1}) + \frac{\partial \mathbf{f}_i}{\partial \mathbf{x}_i}(\mathbf{x}_i - \hat{\mathbf{x}}_i) + \frac{\partial \mathbf{f}_i}{\partial \mathbf{a}}(\mathbf{a} - \hat{\mathbf{a}}_{i-1})$$

De aquí se deriva el *Filtro Extendido de Kalman* en donde

$$\hat{\mathbf{a}}_i = \hat{\mathbf{a}}_{i-1} - \mathbf{K}_i \mathbf{f}_i(\hat{\mathbf{x}}_i, \hat{\mathbf{a}}_i)$$

y la covarianza

$$\mathbf{S}_i = (\mathbf{I} - \mathbf{K}_i \mathbf{M}_i) \mathbf{S}_i$$

donde

$$\mathbf{M}_i = \frac{\partial \mathbf{f}_i}{\partial \mathbf{a}}$$

$$\mathbf{W}_i = \frac{\partial \mathbf{f}_i}{\partial \mathbf{x}} \mathbf{\Lambda}_i \left(\frac{\partial \mathbf{f}_i}{\partial \mathbf{x}} \right)^T$$

y

$$\mathbf{K}_i = \mathbf{S}_{i-1} \mathbf{M}_i^T (\mathbf{W}_i + \mathbf{M}_i \mathbf{S}_{i-1} \mathbf{M}_i^T)^{-1}$$

En el caso de la función que determina la coordenada X a partir de la latitud, longitud y peso esta determinada por:

$$X = (v + h) \cos(\phi) \cos(\lambda)$$

donde

$$v = \frac{a}{\sqrt{1 - e^2 \sin^2(\phi)}}$$

y

$$e^2 = \frac{a^2 - b^2}{a^2}$$

Siendo a el semieje mayor, b el semieje menor, ϕ la latitud, λ la longitud y h el peso. Esto nos da como resultado una función de medida de la forma:

$$f(x, \phi, \lambda, h) = x - (v + h) \cos(\phi) \cos(\lambda)$$

Es claro que $\frac{\partial f}{\partial x} = 1$ por lo que restaría derivar con respecto a las variables del GPS.

$$\begin{aligned}\frac{\partial f}{\partial \phi} &= -1/2 \frac{(a^2-b^2)(\cos(\phi))^2 \cos(\lambda)}{\left(1-\frac{(a^2-b^2)\sin(\phi)}{a^2}\right)^{3/2}} + \left(\frac{a}{\sqrt{1-\frac{(a^2-b^2)\sin(\phi)}{a^2}}} + h\right) \sin(\phi) \cos(\lambda) \\ \frac{\partial f}{\partial \lambda} &= \left(\frac{a}{\sqrt{1-\frac{(a^2-b^2)\sin(\phi)}{a^2}}} + h\right) \cos(\phi) \sin(\lambda) \\ \frac{\partial f}{\partial h} &= -\cos(\phi) \cos(\lambda)\end{aligned}$$

Tomando como vector de observaciones $\mathbf{x} = [\phi \ \lambda \ h]^T$ y vector de estados $\mathbf{a} = [X]$ se tiene que $\frac{\partial f}{\partial \mathbf{x}} = \left[\frac{\partial f}{\partial \phi} \ \frac{\partial f}{\partial \lambda} \ \frac{\partial f}{\partial h}\right]$ y $\frac{\partial f}{\partial \mathbf{a}} = 1$.

Para la siguiente serie de datos se calcula la covarianza

| ϕ | λ | h |
|-------------------|-------------------|--------------------|
| 0.366799740218530 | 1.562997959135845 | 23.199999999999999 |
| 0.366799806785045 | 1.562997959135845 | 23.300000000000000 |
| 0.366800006484590 | 1.562997692869785 | 23.500000000000000 |
| 0.366800206184135 | 1.562997692869785 | 23.800000000000000 |
| 0.366800206184135 | 1.562997692869785 | 23.399999999999999 |
| 0.366800206184135 | 1.562997692869785 | 23.500000000000000 |
| 0.366800206184135 | 1.562997692869785 | 23.399999999999999 |
| 0.366800239466520 | 1.562997692869785 | 23.199999999999999 |
| 0.366800239466520 | 1.562997559736755 | 23.300000000000000 |
| 0.366800239466520 | 1.562997559736755 | 23.399999999999999 |
| 0.366800339317165 | 1.562997559736755 | 23.800000000000000 |
| 0.366800405883680 | 1.562997559736755 | 23.800000000000000 |
| 0.366800439166065 | 1.562997426603725 | 23.899999999999999 |

$$\mathbf{\Lambda} = \begin{bmatrix} 0.000000000000043 & -0.000000000000029 & 0.000000032728551 \\ -0.000000000000029 & 0.000000000000023 & -0.000000023298280 \\ 0.000000032728551 & -0.000000023298280 & 0.060000000000000 \end{bmatrix}$$

Tomando como estimación inicial el cálculo de la primera estimación como datos a prior:

$$\hat{\mathbf{a}}_0 = 46485$$

y

$$\mathbf{S}_0 = 1$$

se tienen las siguientes iteraciones:

| i | $\hat{\mathbf{a}}_i$ | \mathbf{S}_i |
|-----|---------------------------|---------------------------|
| 0 | 46485 | 1 |
| 1 | $4.648594023556208e + 04$ | $3.100409133871551e - 06$ |
| 2 | $4.648594005972359e + 04$ | $1.550204891086269e - 06$ |
| 3 | $4.648646834147919e + 04$ | $1.033493630062156e - 06$ |
| 4 | $4.648673215299317e + 04$ | $7.751291739997000e - 07$ |
| 5 | $4.648688986046493e + 04$ | $6.201076302325359e - 07$ |
| 6 | $4.648699512131807e + 04$ | $5.167587434589532e - 07$ |

Aun cuando la información a prior se encuentre muy lejana al valor verdadero el algoritmo solo requiere de un iteración para obtener el valor ya que únicamente un parámetro se está calculando.

El procedimiento para la estimación de las coordenadas Y y Z se sigue de la misma forma.

Bibliografía

- Agency, N. M. (2010). A guide to coordinate systems in great britain. Technical report, Ordnance Survey.
- Botterill, T., Mills, S., and Green, R. (2010). Real-time Aerial Image Mosaicing. In *Proceedings of Image and Vision Computing New Zealand*, pages 1–6, Queenstown, NZ.
- Braden, B. (1986). The surveyor’s area formula. *The College Mathematics Journal*.
- Capel, D. (2001). *Image mosaicing and super-resolution*. PhD thesis, Oxford.
- Curtis, M. L. (1984). *Matrix Groups*. Springer.
- de Berg, M., Cheong, O., van Kreveld, M., and Overmars, M. (2008). *Computational Geometry: Algorithms and Applications*. Springer.
- Hannuksela, J., Sangi, P., Heikkilä, J., Liu, X., and Doermann, D. S. (2007). Document image mosaicing with mobile phones. In *ICIAP*, pages 575–582. IEEE Computer Society.
- Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Proceedings of the Alvey Vision Conference, University of Manchester*, pages 147–151. The Plessey Company.
- Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN:0521549518.
- Heikkilä, M. and Pietikäinen, M. (2005). An image mosaicing module for wide-area surveillance. In *Proceedings of the third ACM international workshop on Video surveillance & sensor networks, VSSN '05*, pages 11–18, New York, NY, USA. ACM.
- Jensen, A. M., Morgan, D., Clemens, S., Chen, Y., and Hardy, T. (2009). Using multiple open-source low-cost unmanned aerial vehicles (uav) for 3d photogrammetry and distributed wind measurement. In *Proceedings of the 2009 ASME/IEEE International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*.

- Kaplan, E. D., Leva, J. L., Milbert, D., and Pavloff, M. S. (2006). *Understanding GPS*, chapter Fundamentals of Satellite Navigation. Artech House.
- Ma, Y., Soatto, S., Kosecka, J., and Sastry, S. S. (2003). *An Invitation to 3-D Vision*. Springer.
- Margolis, M. (2011). *Arduino Cookbook*. O'Reilly Media.
- Oh, P. Y. and Green, B. (2003). A kite and teleoperated vision system for acquiring aerial images. In *IEEE International Conference on Robotics and Automation (ICRA)*.
- Orr, M. (1992). *A Mickey Mouse guide to Kalman Filtering*. DAI working paper. Edinburgh University.
- Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Halsted Press.
- Slama, C. (1966). *Manual of photogrammetry*. American Society of Photogrammetry.
- Tapp, K. (2000). *Matrix Groups for Undergraduate*. American Mathematical Society.
- Welch, G. and Bishop, G. (1995). An introduction to the kalman filter. Technical report, Chapel Hill, NC, USA.
- Xiong, Y., Wang, X., Tico, M., Liang, C.-K., and Pulli, K. (2009). Panoramic imaging system for mobile devices. In *SIGGRAPH Posters*.
- Xu, G. (2007). *GPS: Theory, Algorithms and Applications*. Springer.

Índice alfabético

Área de traslape, 35

Acelerometro, 9

Análisis de estabilidad, 31

Arduino, 6

Correlación cruzada, 42

DLT, 20

Fórmula de Surveyor, 37

Formula de Rodrigues, 16

Harris, 39

Homografía, 20, 33

Marco de referencia global, 25

Marco de referencia local, 27

Matriz de calibración, 20

Matriz de rotación, 17

Modelo de cámara, 18

Picavet, 10

Proyección, 50

RANSAC, 43

Receptor GPS, 9

Rotación de la cámara, 27

Sistemas de posicionamiento global, 11