

# Adquisición y Procesamiento de Imágenes Aéreas para Sensado Remoto

José F. Manera<sup>1</sup>, Lucas Rodríguez<sup>1</sup>, Claudio Delrieux<sup>2,3</sup> and Ricardo Coppo<sup>2</sup>

<sup>1</sup> Departamento de Ciencias e Ingeniería de la Computación

<sup>2</sup> Departamento de Ingeniería Eléctrica y de Computadoras

<sup>3</sup> IIIE - CONICET, parcialmente financiado por SECyT-UNS

Universidad Nacional del Sur - Alem 1253 - (8000) Bahía Blanca - Argentina

## Resumen

Los sistemas de adquisición de imágenes aéreas digitales de alta resolución basados en cámaras de formato pequeño representan una alternativa de gran versatilidad en la solución de diversos problemas de monitoreo ambiental y sensado remoto, particularmente aquellas obtenidas en base a dispositivos aéreos autónomos UAV. Esto es así dado que un correcto procesamiento de imágenes de cámaras de formato pequeño (ópticas y/o multiespectrales) permite la obtención de ortomosaicos que cumplan con estándares de calidad, a una fracción del costo operativo del uso de equipos aerotransportados de costo mucho mayor, y con mayor flexibilidad y resolución que con el uso de imágenes satelitales. El objetivo de este trabajo de investigación es la implementación de un sistema de cómputo que permita sistematizar la adquisición, procesamiento, rectificación, formación de mosaicos, y georreferenciación de las imágenes adquiridas por medio de cámaras ópticas y multiespectrales transportadas por un UAV. Los resultados obtenidos permiten automatizar el procesamiento casi por completo, requiriendo un mínimo de atención no especializada, y permiten organizar la información e imágenes obtenidas en los vuelos para su posterior uso en procesos de reconocimiento, interpretación e identificación.

## 1. Introducción

La teledetección y el sensado remoto constituyen una de las metodologías preferidas para la obtención de información para el monitoreo ambiental no invasivo [5,6]. Esta tecnología cuenta con un sinnúmero de aplicaciones en diversas áreas de la actividad humana, desde estudios científicos y ecológicos, hasta prospectivas económicas, urbanas y sociales, y por lo tanto constituye también un instrumento específicamente utilizado para la obtención de información cuantitativa y cualitativa para la toma de decisiones y fijación de políticas [4].

### 1.1. Alternativas en la obtención de imágenes de sensado remoto

Las imágenes generadas procesando información de sensado remoto tienen en particular un enorme potencial dentro de todas estas áreas, y han concitado

recientemente una notable atención masiva gracias a aplicativos específicos como por ejemplo el Google Earth. Dentro del diverso espectro de modalidades e instrumentos para el sensado remoto, podemos mencionar las imágenes satelitales de diversas misiones y constelaciones, las imágenes obtenidas por medio de plataformas aerotransportadas, y las imágenes obtenidas por medio de cámaras fijas.

Cada una de dichas modalidades tiene ventajas y desventajas específicas. Por ejemplo, las imágenes satelitales suelen poseer parámetros de calidad insuperables, y entregan información espectacularmente calibrada. Sin embargo, las imágenes satelitales disponibles en forma gratuita son de muy baja resolución espacial, no son obtenibles con la frecuencia necesaria para muchos estudios, y las condiciones climáticas no son negociables (p.ej. presencia de nubes, niebla, u otros factores que afectan las mediciones). Imágenes satelitales de mayor resolución espacial suelen ser prohibitivamente caras.

Las imágenes obtenidas por medio de plataformas aerotransportadas se sobreponen a la falta de versatilidad de las imágenes satelitales, fundamentalmente porque pueden ser obtenidas en condiciones óptimas de clima, iluminación, etc., y la resolución espacial puede acomodarse a los dispositivos sensores, altitud de vuelo, etc. La desventaja fundamental de esta modalidad es que el costo operativo es muy alto, y no se han desarrollado hasta ahora sistemas de procesamiento móviles que permitan verificar en tiempo real si la adquisición se ha realizado en forma adecuada. Las imágenes de cámaras fijas suelen ser de calidad baja o media, y requieren un procesamiento muy complejo para ser de alguna utilidad a la hora de establecer mediciones cuantitativamente precisas. Tienen además la desventaja fundamental de tener un alcance limitado.

## 1.2. Objetivos de este trabajo

Una alternativa reciente, que permite sobreponerse a muchas de las limitaciones mencionadas más arriba, es la utilización de sistemas de adquisición de imágenes aéreas digitales de alta resolución basados en cámaras de formato pequeño obtenidas en base a dispositivos aéreos autónomos UAV. El uso de imágenes de cámaras de formato pequeño (ópticas y/o multiespectrales), con una adecuada calibración, e incorporando información adicional del terreno, permite la obtención de resultados que combinan la calidad de las imágenes de las mejores misiones satelitales disponibles o de imágenes de equipos aerotransportados convencionales, con un costo mucho menor [8].

El objetivo de este trabajo es el desarrollo de un sistema que permita automatizar los diferentes pasos necesarios para el procesamiento de imágenes adquiridas por medio de cámaras ópticas y multiespectrales transportadas por un UAV. Para ello contamos con la colaboración de la empresa NOSTROMO S.A. ([www.nostromo-defensa.com](http://www.nostromo-defensa.com)) quienes han facilitado un conjunto de imágenes de prueba.

El sistema procesa todas las imágenes utilizando la frase GPS presente en los archivos (o la información de GPS obtenida en el logger de la computadora de vuelo) para ensamblarlas en un mosaico. Luego, para los bitmaps vecinos

(con la vecindad determinada por la georreferenciación de cada imagen), localiza un conjunto de características (features) en común, y entre ellas conforma el triángulo de mayor área posible. Por último, dados dos triángulos de imágenes vecinas, encuentra la transformación afín que permite mapear a los vértices del triángulo en una imagen sobre los mismos vértices en la otra. Esa misma transformación se aplica luego a todos los pixels de la imagen, obteniéndose la unión o "stitch" entre ambas. Cuando la precisión de esta operación introduce discontinuidades, es posible aplicar una interpolación bilineal entre ambas imágenes, obteniéndose un suavizado que hace que la unión sea imperceptible. Eventualmente las imágenes así procesadas pueden subirse al Google-Earth para su utilización remota por parte de usuarios autorizados.

Este trabajo está organizado de la siguiente manera: en la siguiente sección se describe la metodología desarrollada para este trabajo, en particular la adquisición y procesamiento de las imágenes y su metadata, y los algoritmos implementados para elaborar los mosaicos. En la sección 3 se muestran los resultados obtenidos, y la sección 4 concluye el trabajo con la discusión y los trabajos futuros.

## 2. Metodología

La posibilidad de incorporar pequeñas plataformas con cámaras y sensores en UAVs (unmanned aircraft vehicles) permite actualmente un amplio abanico de posibilidades de sensado remoto para usos civiles, desde cámaras digitales convencionales (a las cuales se les debe adaptar electrónicamente el disparo del *shutter*), hasta sensores laser, GPS, cámaras infrarrojas, multispectrales, etc. Estos y otros dispositivos, así como los actuadores específicos para la navegación del UAV, son comandados por el autopiloto o computadora de vuelo. El autopiloto se programa en tierra, y adquiere el control del vuelo luego del despegue manual y hasta el momento del aterrizaje.

Dentro de su programación se incluye el plan de vuelo, el cual se posiciona por medio del GPS, así como las localizaciones geográficas y actitudes donde debe ocurrir disparo del *shutter* de las cámaras o sensores. Finalmente, el autopiloto registra en un logger la información eventual que se va registrando durante la misión. Las imágenes adquiridas durante el vuelo, junto con el data logger, quedan registradas en una memoria desde la cual se descargará la información una vez finalizada la misión.

### 2.1. Adquisición y manejo de las imágenes

Las imágenes disponibles para este trabajo, y cuyo procesamiento será ulteriormente utilizado en diversas aplicaciones, son en principio adquiridas desde una cámara multispectral TETRACAM, la cual está específicamente diseñada para su montaje en plataformas ultralivianas aerotransportadas. Para ello cuenta con un *shutter* electrónico, disparado por la computadora de vuelo, el cual también accede al GPS y registra la frase de posicionamiento global y tiempo.

La imagen adquirida, junto con la frase GPS, quedan almacenados en un archivo RAW de formato propietario, en el cual el bitmap queda registrado en el formato Bayer nativo del sensor. Para manipular estas imágenes, se desarrolló un *reader* que decodifica el bitmap del archivo RAW, le aplica una función de deBayering por interpolación lineal, y la lectura de la metadata que incluye la frase del GPS.

También se contó con imágenes digitales de cámaras convencionales, las cuales no registran el posicionamiento. Para poder referenciarlas geográficamente, contamos con el logger de la computadora de vuelo, el cual registra también la frase del GPS.

## 2.2. Armado del mosaico

Si bien existen programas disponibles que permiten armar mosaicos a partir de imágenes digitales (por ejemplo, Autopano, Autostitch, Photostitch, e inclusive funciones específicas del Photoshop), ninguna de estas herramientas se desempeña adecuadamente con las imágenes aéreas con las que contamos. En principio el mosaico de imágenes podría armarse buscando la correspondencia de a pares, eso eventualmente llevaría a un problema de búsqueda de gran complejidad computacional. Por lo tanto, y dado que se cuenta con la información geográfica del punto de adquisición de las imágenes, eso permitiría un pre-ordenamiento de las mismas para facilitar los siguientes pasos de procesamiento.

El armado de un mosaico se basa en el *matching* entre pares de imágenes contiguas. Para ello se requiere encontrar una transformación afin entre ellas que ponga en correspondencia directa a los puntos de una dentro del sistema de referencia de la otra. Este procesamiento puede basarse en dos estrategias [1,7]. La primera de ellas, basada en características (*feature based*), consiste en encontrar puntos característicos que se repitan en el área común entre ambas imágenes. Idealmente se podría encontrar la correspondencia óptima entre cada par de pixels de las imágenes a aparear, es decir, una transformación específica para cada uno. En la práctica se toma un conjunto limitado de puntos destacados, de manera similar al mecanismo de georreferenciación que se utiliza con imágenes satelitales.

La segunda estrategia (*pixel based*) se basa en procesar bloques de pixels entre ambas imágenes, buscando una transformación que minimice la diferencia cuadrática acumulada entre pixels correspondientes. El problema en este caso es que el espacio de búsqueda es muy grande, y el cómputo de la diferencia cuadrática también es costoso.

En nuestra implementación, utilizamos una estrategia mixta, que combina las ventajas de ambas estrategias. En un primer paso se encuentran puntos característicos en correspondencia, con los cuales se determina una primera transformación afin, la cual es tomada como una primera aproximación. Luego, se aplica un algoritmo de minimización de la diferencia cuadrática, buscando la transformación que optimiza dicha diferencia en un entorno alrededor de esta aproximación.

Los puntos característicos se detectan por medio del algoritmo SURF (Speeded Up Robust Features) [2], empleando una implementación OpenSource del algoritmo. Este algoritmo evalúa un conjunto de descriptores de imagen (features) para los píxeles de una imagen, y determina un subconjunto de píxeles característicos por el hecho de poseer valores destacados en algunos de dichos descriptores.

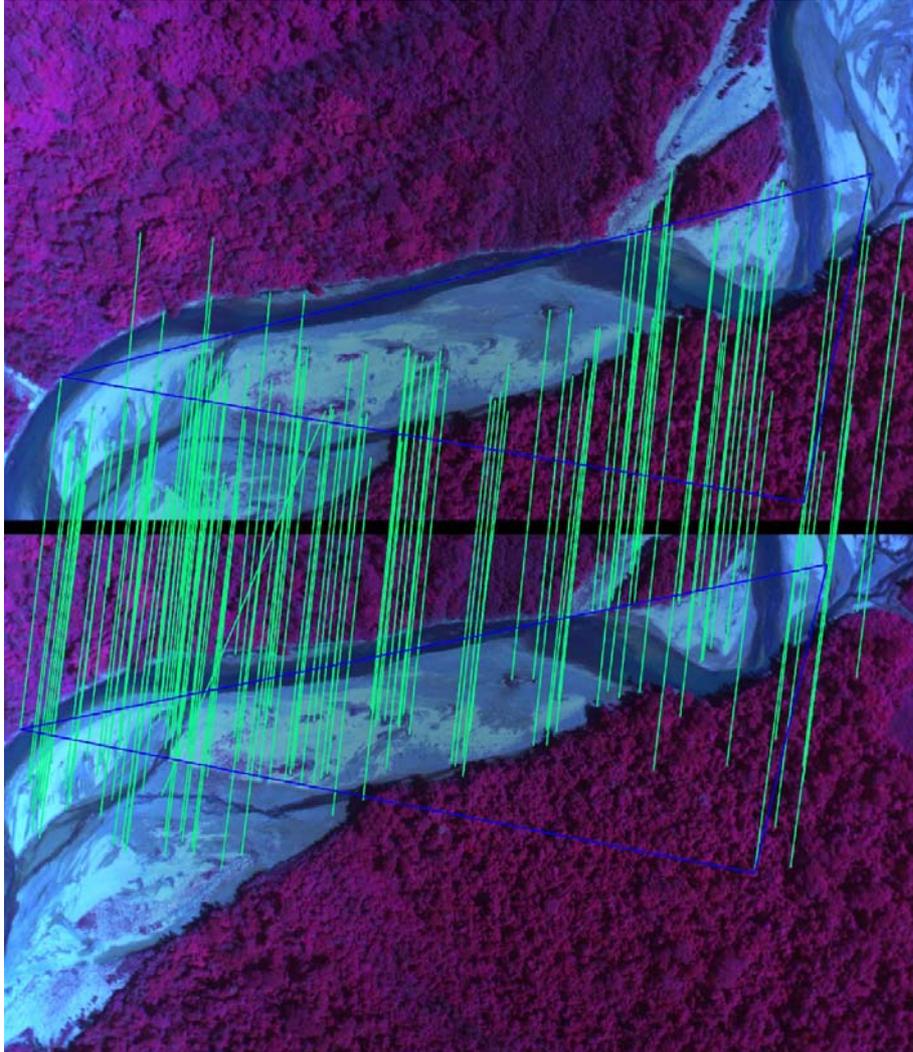
Para encontrar la correspondencia entre puntos característicos en dos imágenes a aparear, se puede tener en cuenta la posición geográfica de los mismos a partir de la información del GPS y la resolución espacial de las imágenes. Cuando este criterio no es lo suficientemente preciso (por ejemplo, hay varios puntos característicos muy próximos entre sí), se adopta un segundo criterio que se basa en comparar el vector de descriptores de los puntos característicos. Para ello, se buscan los pares de puntos característicos más próximos en el espacio de los descriptores. En este caso, la noción de proximidad en el espacio de descriptores que se utilizó fue la distancia de Manhattan (es decir, la sumatoria de la diferencia absoluta entre pares de valores de parámetros). Este segundo criterio resultó importante a la hora de reducir el espacio combinatorio de búsqueda entre pares de puntos destacados correspondientes entre ambas imágenes.

Una vez encontrado un conjunto de pares de puntos correspondientes entre las imágenes a aparear, se busca entre ellos el triángulo de mayor área para determinar la transformación afín que ponga en correspondencia a todos los píxeles de la imagen. Este criterio busca reducir el efecto que tendría elegir tres puntos muy próximos para determinar los parámetros de la transformación afín, dado que, en ese caso, el posible error geométrico en la determinación numérica de los parámetros de la misma se vería amplificado en las zonas más alejadas de la imagen. En la Fig. 1 se puede observar un par de imágenes a aparear, los pares de puntos característicos correspondientes en una y otra (unidos en verde), y el triángulo de área máxima (en azul).

A partir de la determinación de la transformación *feature based* óptima, se utiliza el método *pixel based* para buscar localmente la transformación que minimice el error cuadrático acumulado en la zona de superposición entre las imágenes a aparear. Para suavizar el efecto producido por la unión entre las imágenes, si es necesario se aplica interpolación lineal (*blending*) entre ellas. En la Fig. 2 se puede ver la ampliación de una zona con la unión entre tres imágenes, comparando la mejora que se produce al utilizar *blending*.

### 3. Resultados

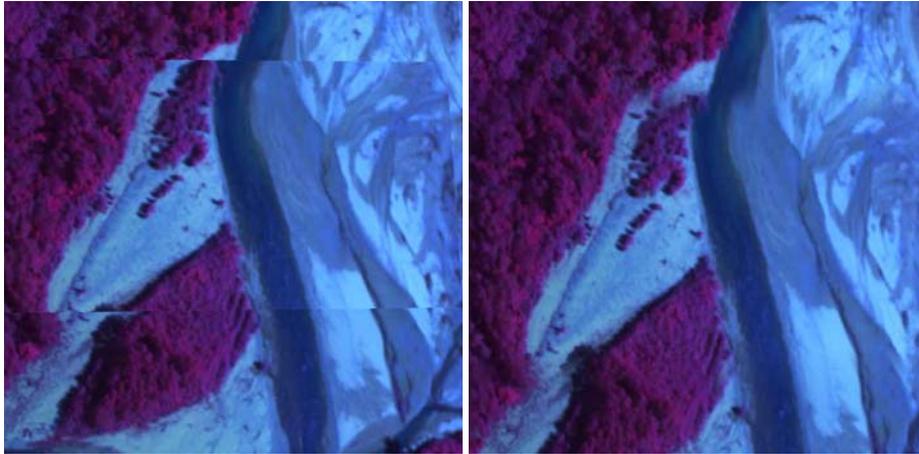
Las ideas mencionadas en la sección anterior fueron implementadas y testeadas en pares de imágenes hasta lograr optimizar los resultados. También se hicieron pruebas con imágenes aéreas de cámaras digitales estándar, las cuales son en general de mala calidad dado que el enfoque no es constante, generan el efecto almohadilla por el cual la imagen se deforma y pierde gradualmente luminosidad hacia los rincones. Los resultados con este tipo de imágenes, sin embargo, son bastante aceptables (ver Fig. 3). Con imágenes de muy alta calidad, como



**Figura 1.** Dos imágenes a aparear, los pares de puntos característicos correspondientes en una y otra (unidos en verde), y el triángulo de área máxima (en azul).

por ejemplo pruebas realizadas con recortes de imágenes IKONOS, los resultados son excelentes (ver Fig. 4).

De acuerdo a las pruebas realizadas, nuestro sistema está en condiciones de tomar un conjunto de imágenes TETRACAM, y en forma automática determinar su posición accediendo a la localización geográfica registrada en la frase GPS contenida en la metadata, convertir el formato RAW a mapa de bits utilizando de Bayering lineal, y armar el mosaico uniendo los pares de imágenes contiguas



**Figura 2.** Ampliación de la unión entre tres imágenes: (a) directa, y (b) con blending.

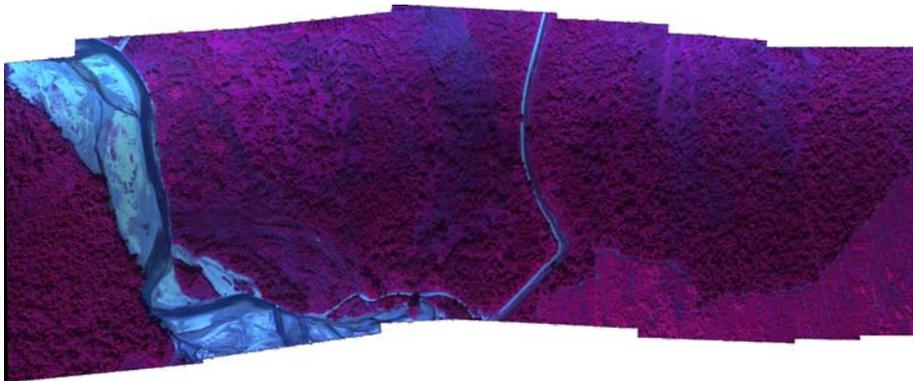


**Figura 3.** Unión entre dos imágenes aéreas de cámaras digitales estándar.



**Figura 4.** Unión entre recortes de una imagen IKONOS.

aplicando el algoritmo mencionado. Un ejemplo de este resultado puede apreciarse en la Fig. 5.



**Figura 5.** Mosaico automáticamente generado a partir de varios archivos TETRA-CAM.

Los mosaicos así obtenidos pueden ser objeto de ulterior procesamiento, según los pasos típicos del tratamiento digital de imágenes, para la obtención de información de sensado remoto que pueda ser considerada valiosa en las aplicaciones específicas. Al mismo tiempo, al tratarse de mosaicos que pueden ser fácilmente georreferenciados, es posible generar layers de información accesibles desde Google Earth para su uso interactivo.

#### **4. Conclusiones y Trabajo Futuro**

Se desarrolló de un sistema que permite llevar adelante los diferentes pasos necesarios para el procesamiento de imágenes adquiridas por medio de cámaras

ópticas y multispectrales transportadas por un UAV, y el armado de mosaicos a partir de las mismas, de manera automática. El sistema utiliza la metadata GPS presente en los archivos (o la información de GPS obtenida en el logger de la computadora de vuelo) para determinar cuáles imágenes deberían ser contiguas en el mosaico, convierte los datos RAW de los archivos en bitmap (aplicando de-Bayering), y para los bitmaps contiguos localiza un conjunto de características en común, y entre ellas conforma el triángulo de mayor área posible. Finalmente, dados dos triángulos de imágenes vecinas, se encuentra la transformación afín que permite mapearlos, en una imagen, sobre los mismos vértices en la otra. Dicha transformación es luego refinada por un algoritmo local que minimiza la diferencia cuadrática entre acumulada entre los pixels de la unión. La transformación refinada se aplica luego a todos los pixels de la imagen, obteniéndose una unión entre ellas. Cuando la precisión de esta operación introduce discontinuidades, es posible aplicar una interpolación bilineal entre ambas imágenes, obteniéndose un suavizado que hace que la unión se haga menos perceptible. Eventualmente las imágenes así procesadas pueden subirse al Google Earth para su utilización remota por parte de usuarios autorizados.

Los resultados de este trabajo constituyen en realidad la base a partir de la cual se desarrollarán diferentes aplicaciones de uso civil, en particular el monitoreo de cultivos, la evaluación de siniestros, estudios forestales, y agricultura de precisión. Todas esas aplicaciones demandarán la implementación de algoritmos específicamente adaptados para este tipo de imágenes [3].

## Referencias

1. A. Agarwala. Interactive digital photomontage. *ACM Transaction on Graphics*, 23(3):292–300, 2004.
2. Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding*, 110(3):346–359, 2008.
3. J.R. Jensen. *Digital Image Processing: a Remote Sensing Perspective*. Prentice-Hall, Cambridge, third edition, 2005.
4. J.R. Jensen. *Remote sensing of the environment: an Earth resource perspective*. Prentice-Hall, Cambridge, second edition, 2007.
5. T. Lillesand and R. Kiefer. *Remote Sensing and Image Interpretation (4th. ed.)*. Willey & Sons, New York, 2000.
6. Robert Schowengerdt. *Remote Sensing Models and Methods for Image Processing (2th. ed.)*. Academic Press, San Diego, 1997.
7. R. Szeliski. Image alignment and stitching: A tutorial. Report MSR-TR-2004-92, Microsoft Research, 2004.
8. G. Zhou, V. Ambrosia, A.J. Gasiewski, and G. Bland. Foreword to the Special Issue on Unmanned Airborne Vehicle (UAV) Sensing Systems for Earth Observations. *Geography & Remote Sensing*, 47(3):687–689, 2009.