



## Construcción automática de ortofotomapas: una aproximación fotométrica

R. Prados<sup>a,\*</sup>, R. García<sup>a</sup>, L. Neumann<sup>b</sup>

<sup>a</sup>Computer Vision and Robotics Group, University of Girona, 17071 Spain

<sup>b</sup>Computer Vision and Robotics Group, University of Girona, 17071 Spain, ICREA, Barcelona, 08010 Spain

### Resumen

La construcción de mosaicos de imágenes permite obtener representaciones de grandes dimensiones y resolución de una misma escena. Son frecuentes hoy día las cámaras fotográficas que incorporan un software destinado a su construcción o aplicaciones en línea como Google Maps que permiten visualizar mapas resultantes de la construcción de foto-mosaicos. Habitualmente los mosaicos panorámicos son generados a partir de imágenes adquiridas mediante una cámara que únicamente efectúa movimientos de rotación alrededor de un punto fijo. Cuando las condiciones de adquisición varían y la cámara también se traslada, surgen fenómenos, como el de paralaje, que dificultan la unión no perceptible de las imágenes. A ello hay que añadir las diferencias en apariencia que varias fotografías adyacentes pueden presentar debido a mecanismos automáticos de las cámaras, como el de control de exposición. En el presente trabajo se describe un procedimiento completo para la construcción automática de mosaicos con apariencia totalmente continua y consistente, en los que las uniones de las distintas imágenes que lo conforman no son visibles. Las imágenes son registradas mediante métodos que garantizan consistencia geométrica, y unidas utilizando técnicas de fusión (o *blending*), con el objetivo de asegurar una transición no visible entre imágenes y una apariencia global coherente en todo el mosaico. El procedimiento descrito es aplicado sobre una secuencia con el fin de evaluar su utilización en el contexto de las imágenes aéreas de grandes dimensiones. Copyright © 2013 CEA. Publicado por Elsevier España, S.L. Todos los derechos reservados.

### Palabras Clave:

Procesamiento de imagen, realzado de imagen, emparejamiento de imágenes, registro de imágenes, métodos de gradiente

### 1. Introducción

La unión de diversas imágenes de una misma escena en una única imagen de mayores dimensiones se conoce como *Mosaico*. La construcción de mosaicos es muy común en la actualidad, y son numerosas las cámaras fotográficas que incorporan un software con la opción de generarlos a partir de una secuencia de imágenes. En el caso concreto de los mosaicos panorámicos, la cámara se suele colocar en un trípode, o se mantiene lo más fija posible en una posición, de forma que sólo pueda efectuar movimientos de rotación, aproximadamente sobre el punto focal de la cámara.

Sin embargo, en cuanto las condiciones de adquisición difieren del caso descrito, que se considera ideal, pueden darse circunstancias que dificulten su generación. Así, desde un punto de vista geométrico, si la cámara no realiza únicamente un movimiento de rotación, sino que también hay de traslación, aparece el efecto conocido como paralaje. Dicho fenómeno consiste

en la desviación angular de la posición aparente de un objeto, producida por la variación del punto de vista, y se traduce en inconsistencias en el momento de superponer dos imágenes adyacentes. Dichas inconsistencias se acentúan a medida que la distancia entre la cámara y la escena decrece.

A menudo, las cámaras son utilizadas con el mecanismo de cálculo de exposición automática activado. Desafortunadamente, dicha exposición automática puede dar lugar a problemas, como el de la disparidad en la apariencia de imágenes geométricamente adyacentes (ver Figura 1). Si a ello añadimos la posibilidad de unir fotografías adquiridas en días distintos, o bajo condiciones de iluminación también cambiantes, la visibilidad de la unión entre dichas imágenes puede resultar más que evidente.

La construcción de mosaicos consta de tres etapas: registro geométrico de las imágenes, procesado fotométrico, y finalmente una etapa de fusión (o *blending*). El objetivo de la etapa final es obtener una transición suave y sin artefactos entre las diferentes imágenes que componen el mosaico, de manera que el resultado sea lo más homogéneo y convincente posible.

El principal problema que se desea abordar en el presente trabajo es el de la construcción automatizada de ortofotomoto-

\*Autor en correspondencia

Correos electrónicos: [rprados@eia.udg.edu](mailto:rprados@eia.udg.edu) (R. Prados), [rafael.garcia@eia.udg.edu](mailto:rafael.garcia@eia.udg.edu) (R. García), [lneumann@eia.udg.edu](mailto:lneumann@eia.udg.edu) (L. Neumann)

saicos, mediante la unión de una secuencia de imágenes aéreas de alta resolución. El ortofotomosaico resultante debe presentar una apariencia uniforme, y las uniones entre las imágenes que lo conforman deben ser imperceptibles.

La importancia de una correcta y convincente unión de un conjunto de imágenes de una misma escena va más allá de la mera obtención de un buen aspecto estético. Para los científicos expertos, pertenecientes a los campos de la arqueología, biología (Naydenova and Jelev (2009)) o geología (Escartín et al. (2008)), resulta mucho más cómodo y eficiente poder trabajar con una única imagen, de grandes dimensiones, continua y coherente, que no hacerlo con un conjunto de imágenes aisladas y con diferencias radiométricas entre ellas. Así, por ejemplo, Solberg et al. (1996) evaluaron positivamente los beneficios de la utilización de imágenes de satélite de alta resolución para la generación de mapas y clasificación de especies de árboles en los bosques, o Lear (1997) apuntó el interés de diversos campos, como el estudio del suelo o de la agricultura, por la utilización de dicho tipo de imágenes de grandes dimensiones.

Uno de los principales ejemplos de generación de ortofotomapas de grandes dimensiones es Google Maps. Los mapas ofrecidos por dicho servicio, que gozan de una resolución más que aceptable en función de la región geográfica cubierta, presentan una apariencia consistente en áreas relativamente extensas. Sin embargo, son evidentes las transiciones entre transectos pertenecientes a momentos de adquisición diferentes, con condiciones ambientales cambiantes o capturados con una cámara con distintas características y/o configuración (ver Figura 1).



Figura 1: Ejemplo de artefactos en la transición entre imágenes presentes en Google Maps (afuera de la ciudad de Girona). Los artefactos son debidos no únicamente a las diferencias en apariencia de las imágenes implicadas sino también a un error en el alineamiento geométrico de las mismas.

En el presente trabajo se describen las fases de procesamiento de una secuencia de imágenes, correspondientes al lago de Banyoles (en la provincia de Girona, España), capturadas con una cámara de 39 Mpíxeles desde un ultraligero. La aplicación de dicho mosaico será el estudio y control de todo el perfil del lago desde los puntos de vista geológico y biológico.

### 1.1. Secuencia de imágenes

La secuencia utilizada para la construcción del mosaico consta de 41 imágenes, adquiridas con una cámara profesional Hasselblad H3D-39, equipada con un sensor de 39 Mpíxeles de



Figura 2: Selección de imágenes colindantes de la secuencia procesada. Algunas de las imágenes presentan oleaje en la superficie del lago (derecha e izquierda) mientras que en otras el agua se encuentra en calma (centro).

$36,7 \times 49,00$  mm. Cada una de las imágenes tiene una resolución de  $6216 \times 5412$  píxeles, con una profundidad de color de 16 bits por canal, y un total de tres canales, al tratarse de imágenes en color. Dicha información se encuentra almacenada en formato RAW, y en canales separados, aunque también se dispone de la correspondiente versión en formato JPG, para una rápida manipulación y visualización de las mismas.

Así mismo, para algunas de las imágenes se dispone de información de georeferenciación. Concretamente, se conocen las coordenadas UTM de algunos puntos fiduciales, que permiten la construcción de una primera estimación del mosaico. Los puntos, no obstante, son fruto de la marca manual de los mismos, a partir del conocimiento de las coordenadas UTM de dos puntos geodésicos que aparecen en el mapa, por lo que un mosaico basado únicamente en dicha información no sería lo bastante preciso.

Las imágenes fueron adquiridas desde una cámara montada en un aeroplano y orientada hacia abajo. La trayectoria recorrida por el aeroplano obedece a un barrido con transectos Norte-Sur, y se prolongó durante 12 minutos, en una misma jornada, con condiciones lumínicas cambiantes y con alteraciones constantes en la dirección del viento. Debido a dichos cambios de dirección, así como a los pequeños cambios de orientación del aeroplano durante la adquisición, la superficie del lago muestra en algunas de las imágenes ciertos artefactos lumínicos. Los artefactos son debidos a la diferente dirección de incidencia de los rayos solares sobre las pequeñas ondulaciones del agua, y ello provoca que la superficie presente en ocasiones un aspecto significativamente diferente al del resto de imágenes.

## 2. Trabajos previos

El problema de la generación automática de ortofotomapas con apariencia uniforme y verosímil consta de varias etapas, que deben ser completadas de forma secuencial. La literatura al respecto es extensa, aunque no todas las propuestas ponen el mismo énfasis en cada una de las etapas en cuestión. La generación de ortofotomapas se podría considerar incluida dentro de la categoría de los métodos de generación de panoramas, aunque salvando algunas diferencias. En el caso de los panoramas de imágenes, la cámara no acostumbra a moverse significativamente de un punto de rotación, evitándose así efectos como el de paralaje, mientras que en el caso de los ortofotomapas basados en imágenes aéreas el punto de vista cambia constantemente, ya que el aeroplano se desplaza. Por otro lado, mientras que las imágenes que constituyen un panorama suelen ser

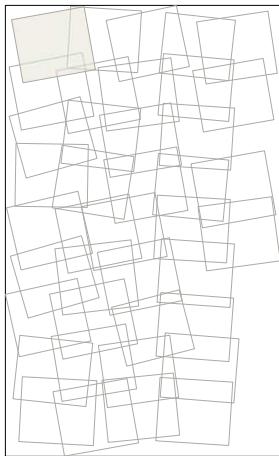


Figura 3: Cobertura de las imágenes adquiridas por el aeroplano siguiendo una trayectoria de barrido de ida y vuelta en dirección vertical, donde el rectángulo sombreado simboliza la primera imagen de la secuencia.

capturadas en un intervalo de tiempo relativamente reducido, las adquisiciones aéreas pueden extenderse durante días, hecho que puede acrecentar a su vez la disparidad de apariencias entre los diferentes transectos de las mismas.

En lo que se refiere estrictamente al registro geométrico, la aproximación más utilizada en la literatura es la basada en búsqueda de correspondencias (Zitová and Flusser (2003)). Dadas dos imágenes, en cada una de ellas se determina un conjunto de puntos de interés, mediante algoritmos como los de Harris (Harris and Stephens (1988)), SIFT (Lowe (2004)) o SURF (Bay et al. (2006)), entre otros. Para dichos puntos su correspondencia en la otra imagen es determinada, tal y como se realiza en trabajos como González et al. (2001). Las correspondencias son utilizadas para el cálculo de una matriz de transformación plana, denominada homografía (Kanatani and Ohta (1999)), que permite convertir las posiciones de los píxeles de una de las imágenes al sistema de coordenadas de la otra. Dicha forma de proceder es factible cuando la escena se puede considerar plana, es decir, cuando el relieve de la escena no es significante respecto de la distancia entre la cámara y la escena. Cuando la escena no obedece a dichas condiciones, es posible recurrir a técnicas de registro no rígido (Arévalo and Gonzalez (2008)), que permiten transformaciones elásticas de las imágenes, que no pueden ser modeladas por una homografía.

Uno de los primeros métodos para el registro de imágenes adquiridas por un vehículo aéreo, con información de georeferencia precisa, fue el propuesto en Kumar et al. (1998), que utilizaba como mecanismo de búsqueda de puntos de interés entre las imágenes un método de filtrado multiescalas, con selección de puntos prominentes (Lindeberg (1993)). Una de las limitaciones del algoritmo se encuentra en el hecho que las relaciones de movimiento entre las imágenes a registrar y el sistema de referencia deben ser mayoritariamente translacionales, con pocos cambios en rotación, escala o oblicuidad. Así mismo, los cambios en la iluminación global debían ser corregidos mediante el filtrado de las imágenes. Dicho método, a su vez, extendía un trabajo previo, concretamente el de Kumar et al. (1995), en

el que se desarrollaba un sistema completo para la generación de mosaicos a partir de imágenes de vídeo aéreas. Dicho trabajo utilizaba el flujo óptico (Horn and Shunck (1981)) para la estimación del movimiento de la cámara. Sin embargo, los métodos basados en el flujo óptico (Girosi et al. (1989)) suelen requerir poco movimiento entre las imágenes capturadas, para que la estimación del mismo sea posible. Ese requisito impide su aplicación sobre secuencias de fotografías, y restringe su uso a secuencias de vídeo con alta frecuencia de imágenes.

Tegolo and Valenti (2001) presentaron un método de generación de mapas caracterizado por la suposición de ausencia de distorsiones y perspectiva en las imágenes, debido a la gran distancia entre el terreno y el vehículo aéreo de adquisición, así como también por la utilización de coordenadas polares para la representación de la información. El método se centra en la concatenación de imágenes consecutivas, realizando un registro basado en el conocimiento de ciertos puntos característicos en las imágenes, pero no aporta soluciones para la corrección de los artefactos en las uniones, ni trata el problema del cierre de bucles. Su aplicación queda restringida, por lo tanto, a un número de escenarios muy reducido.

Son diversos los métodos de la literatura que combinan los ortofotomapas con modelos digitales de elevación, ya sean construidos por medio de otro tipo de sensorización (Lear (1997) o Morse et al. (2010)), como a partir de técnicas de recuperación de estructura a partir del movimiento (o SFM, del inglés *Structure From Motion*) de la cámara (Nicosevici et al. (2009)).

En lo que se refiere al segundo grupo, Wehn et al. (2002) propusieron un sistema completo que abarcaba el hardware de adquisición, el de preprocesado de las imágenes adquiridas y la generación de un ortofotomap, que era proyectado sobre una estimación de la estructura tridimensional del suelo. El algoritmo tiene presente las necesidades de calibración radiométrica, corrección atmosférica y la eliminación de efectos producidos por los reflejos de la luz desde diferentes ángulos en la escena. La etapa de preprocesado mencionada requiere la intervención del usuario. En cuanto a la estimación de la estructura tridimensional de la escena, se utiliza una pirámide multiresolución para la búsqueda de los puntos de interés. Sin embargo, más allá del preprocesado de las imágenes, no se menciona ningún otro mecanismo que asegure una transición no perceptible entre las mismas. En el mismo contexto, Zebedin et al. (2006) utilizaron tanto información multiespectral como técnicas de reconstrucción 3D para la generación de modelos digitales de superficie densos. Así mismo, los modelos son utilizados con fines de clasificación. Dicho tipo de métodos, entre los que también cabría mencionar el trabajo de Ding et al. (2008), pueden llegar a obtener resultados de apariencia convincente. No obstante, la densidad y la precisión de las reconstrucciones no permiten dar a una ortofoto generada, a partir de dicho modelo, la nitidez que las imágenes originales poseían, especialmente si eran de alta resolución. Los efectos de diferentes exposiciones no son compensados, y aunque en visualizaciones de regiones concretas no sean perceptibles, el mapa global sí que presentará dichos artefactos.

Botterill et al. (2010) desarrollaron un *framework* completo para la construcción de mosaicos de imágenes aéreas en tiempo

real. Dicho *framework* utiliza un algoritmo basado en modelos Bag-of-Words (BoW) (como en Nicosevici and Garcia (2009)) para la localización de las imágenes que puedan ser alineadas con la más recientemente adquirida. Los puntos de interés de las imágenes son localizados mediante la aplicación del detector de esquinas de Harris (Harris and Stephens (1988)) o SIFT (Lowe (1999)). Las transformaciones de perspectiva son calculadas utilizando las correspondencias proporcionadas por el BoW y el *framework* BaySAC (Botterill et al. (2009)), que permite la supresión de datos espurios (*outliers*). Finalmente, la costura óptima, es decir, el camino de corte entre dos imágenes que minimiza las diferencias fotométricas entre ambos lados del mismo, es calculada mediante la aplicación del algoritmo de Dijkstra (1959) sobre las diferencias entre las imágenes superpuestas. La aproximación presenta algunas carencias, como son algún mecanismo para la compensación de los diferentes tiempos de exposición, o la aplicación de algún método para evitar la degeneración del aspecto global del mosaico, a medida que el número de imágenes registradas aumenta. El algoritmo óptimo de búsqueda de costuras no asegura tampoco una transición suave entre las imágenes, en caso de imprecisiones en el registro, cuyo efecto, de hecho, puede irse incrementando con el paso del tiempo, debido a la naturaleza secuencial del proceso.

De los métodos mencionados hasta el momento, orientados hacia la generación de ortofotomapas aéreos, sólo algunos de ellos contemplan la necesidad de correcciones en las imágenes para la compensación de efectos lumínicos no deseados, y ninguno de ellos plantea la necesidad de algoritmos para el suavizado de las transiciones.

Uno de los primeros autores en abordar el problema de la generación de mosaicos aéreos (aunque obviando etapas como la de registro), sin artefactos en la transición de las imágenes, fue Milgram (1975), quien desarrolló un algoritmo para la unión de dos imágenes alineadas de forma manual. El método propuso un ajuste de “orden cero” para compensar las diferencias de iluminación, y una estrategia no óptima para la definición de la línea de costura (o *seam*) en la región de superposición. Dicha estrategia buscaba, para cada línea de la imagen, la pareja de píxeles horizontalmente vecinos que ofrecía una transición más suave, es decir, una menor diferencia fotométrica. El método fue ideado para trabajar sobre imágenes entre las que hubiese una traslación horizontal. Alrededor de cada una de esas parejas de píxeles se define una extensión de un número arbitrario de píxeles, con el objetivo de aplicar sobre la misma un promedio ponderado, que permita suavizar la transición. El algoritmo presentado es simple, pero estableció las bases de la mayor parte de los métodos de suavizado que aparecerían en las siguientes décadas. Milgram (1977) propuso posteriormente una mejora para el algoritmo, que añadía un criterio para la selección de los píxeles en la etapa de compensación de las diferencias de iluminación, con tal de tratar correctamente las sombras y los objetos en movimiento, teniendo en cuenta únicamente los niveles de gris más informativos (evitando nubes y otros fenómenos atmosféricos). También fue incluida una función de coste en el cálculo de la costura entre las imágenes, para controlar las coordenadas de los puntos origen y destino.

En el mismo contexto, Zhu and Yang (2000) propusieron hasta tres métodos diferentes, aplicables a la unión sin costuras de imágenes en la generación de ortofotomapas. El primero de ellos es la corrección de los niveles de gris (en el caso de imágenes en escala de grises), de acuerdo con la diferencia en los niveles de luminancia de dos píxeles correspondientes a una misma posición. El segundo consiste en la aplicación de un suavizado o *feathering* a lo largo de una cierta extensión alrededor de la línea de unión. Los métodos de *feathering* permiten suavizar la transición alrededor de una costura, mediante la aplicación de un promedio ponderado de los valores del área de superposición. Finalmente, el tercer método consiste en un suavizado en el dominio *wavelet*. Sin embargo, los primeros dos métodos no tienen la capacidad de tratar con problemas de registro, ni de compensar diferencias de exposición entre imágenes, mientras que el tercero resulta muy costoso en cuanto a recursos computacionales se refiere, condicionando así su aplicación sobre imágenes de grandes dimensiones.

Wen and Zhou (2008) propusieron un método para el cálculo de la costura entre imágenes con superposición. El algoritmo requiere la conversión de las imágenes adquiridas en color al espacio HIS (*Hue, Intensity, Saturation*, i.e., Tono, Intensidad, Saturación), y busca el camino que da lugar a un error mínimo, en cuanto a la suma de diferencias entre píxeles vecinos en los diferentes planos HIS. Así mismo, propone la aplicación de *feathering* sobre un entorno de tamaño adaptativo, para obtener una transición suave entre ambos lados del camino encontrado. El método es capaz de trabajar con imágenes en color, pero su aplicación se centra en el proceso de imágenes registradas horizontal o verticalmente, que no den lugar a trayectorias complejas para la realización del corte. Por otro lado, los problemas relativos a diferentes tiempos de exposición no se contemplan.

A pesar del número relativamente reducido de propuestas en el campo aéreo, en la bibliografía relacionada con la construcción de mosaicos de imágenes aparecen también numerosos métodos para la combinación de varias imágenes, de forma que las costuras entre ellas no sean visibles. Dichos métodos, no obstante, pueden ser también aplicados, dependiendo de la situación, a imágenes aéreas.

La limitación de Milgram (1975) y Milgram (1977) en el número de imágenes a fusionar –restringido a dos– fue abordada, en primera instancia, por Peleg (1981), quien introdujo el concepto de Función de Eliminación de Bordes (SEF, *Seam-Eliminating Function*). SEF se basa en una función de suavizado de la luminancia (i.e., un mapa de ponderación), calculada mediante un algoritmo iterativo de relajación, y que es usada para ajustar la transición entre un número arbitrario de imágenes con superposición (a pesar de que la información de dicha superposición no se utiliza), estableciendo como cero el valor de las diferencias de luminancia alrededor de los bordes.

Burt and Adelson (1983) introdujeron en 1983 el concepto de *Image Spline*, una transformación que permitía obtener una transición suave entre varias imágenes. El método podía ser aplicado en varios escenarios, a diferencia de Milgram (1975), donde el autor se había centrado únicamente en imágenes de satélite. Las imágenes a ser combinadas son descompuestas en un conjunto de componentes pasa-banda, y se aplica sobre ca-

da una de las bandas una *spline* diferente, con una apropiada amplitud de transición. El objetivo es el de fusionar las características con la misma escala en cada uno de los niveles de la descomposición. Finalmente, los componentes procesados son recombinados en la imagen mosaico, mediante una simple adición. El método suprime la visibilidad de las costuras y reduce los efectos de los posibles desalineamientos cuando el registro no es lo bastante preciso, apareciendo como contrapartida dobles contornos o efectos fantasma cuando dicho desalineamiento es significativo.

Pérez et al. (2003) propusieron un método genérico de interpolación, basado en la resolución de las ecuaciones de Poisson, para la edición sin visibilidad en la transición y la clonación de regiones seleccionadas, aplicable en la construcción genérica de mosaicos y ortofotomapas. El método permite suprimir la visibilidad de las uniones a lo largo de las regiones colindantes a los bordes. Así mismo, se basa en la idea de que, mediante una adecuada mezcla de los gradientes de una imagen con los de otra, es posible fusionar de manera convincente las regiones de la imagen (que también pueden ser objetos manualmente seleccionados) con una apariencia transparente. El entorno de trabajo se basa en la ecuación diferencial parcial con condiciones de borde de tipo Dirichlet, las cuales especifican el Laplaciano de una función desconocida sobre un dominio de interés, junto con los valores desconocidos de la función en los límites del dominio. Como extensión de la técnica presentada por Bertalmio et al. (2000), Pérez et al. (2003) propusieron modificar el problema de la interpolación de imágenes, a través de las ecuaciones de Poisson, introduciendo más restricciones en forma de campo de vectores guía. Dicho campo de vectores, que pueden ser o no un campo de gradiente, permite controlar el comportamiento de la interpolación. En el mismo contexto, Levin et al. (2004) propusieron un método basado en varias funciones de coste, para la evaluación de la calidad de la unión definida en el dominio del gradiente. Los autores llamaron GIST (*Gradient-domain Image Stitching*) al entorno de trabajo desarrollado, proporcionando dos métodos principales para la fusión de imágenes. En primer lugar, las imágenes son combinadas en el dominio del gradiente, reduciendo así las inconsistencias globales entre las partes unidas, debidas a cambios en la iluminación y a variaciones en la respuesta fotométrica de la cámara. En una segunda etapa, la imagen de mosaico es inferida mediante optimización sobre la imagen de gradiente, reduciéndose los artefactos alrededor de los límites de las imágenes y la duplicación de bordes.

El problema de la unión de imágenes en tiempo real, para la generación de mosaicos en línea, fue abordado por Zhao (2006). El autor propuso un método eficiente de fusión para la creación de imágenes de mosaico dinámicas de alta calidad y con generación en tiempo real, a partir de un número arbitrario de imágenes. Desafortunadamente, el método presenta algunas desventajas, en ciertas circunstancias, respecto a los métodos convencionales. En primer lugar, la etapa de fusión se basa en una versión mejorada de un promedio ponderado multiresolución Burt and Adelson (1983), que prioriza los píxeles cercanos al centro de las imágenes, pero dicha estrategia no siempre ofrece los mejores resultados cuando aparecen problemas de registro. En segundo lugar, el mecanismo de corrección

de la exposición toma como referencia el propio mosaico generado hasta el momento. Este hecho puede conducir a una degeneración de la exposición global, cuando alguna de las imágenes implicadas presenta una excesiva o escasa exposición.

Más recientemente, Mills and Dudek (2009) presentaron una combinación de técnicas para la creación de mosaicos de imágenes de alta calidad a pesar de la presencia de objetos en movimiento en la escena. La técnica utiliza funciones heurísticas para determinar la costura óptima, tanto en los dominios de la luminancia como del gradiente, de manera combinada con un algoritmo de splining multiresolución Burt and Adelson (1983), que permite refinar los resultados alrededor de los bordes seleccionados. La compensación de exposición de las imágenes se lleva a cabo basándose en el mosaico generado hasta el momento, aunque dicho método puede conducir a una degeneración del mismo, a medida que se incrementa el número de imágenes. Además, el algoritmo de fusión utilizado por el método puede producir dobles contornos, especialmente en presencia de estructuras complejas en la escena.

El método propuesto se basa en varios de los conceptos mencionados en el resumen de trabajos previos relacionados, intentando aplicar las técnicas más adecuadas para la resolución del problema que nos ocupa.

### 3. Construcción del mosaico

La construcción del mosaico se divide en diferentes etapas ejecutadas de manera secuencial. Dichas etapas incluyen la generación, en primer lugar, de una primera estimación geométrica del mosaico, basada en los datos de georeferenciación marcados manualmente en todas las imágenes de la secuencia. Una vez la topología de la trayectoria ha sido estimada (ver Figura 3), y por lo tanto las superposiciones entre imágenes pertenecientes a los distintos transectos han sido determinadas (Elibol et al. (2010)), es posible proceder a una búsqueda de puntos y correspondencias entre imágenes consecutivas y no consecutivas con superposición de áreas comunes. Dichas correspondencias serán utilizadas, a su vez, y en combinación con los puntos fiduciales marcados manualmente, para refinar el registro geométrico de las imágenes, mediante la aplicación de un algoritmo de alineamiento global, tal y como se describe en Ferrer et al. (2007). Dado que las imágenes pueden haberse visto afectadas por los diferentes mecanismos de ajuste automáticos de la cámara, como la apertura o el tiempo de exposición, es necesario efectuar una ecualización fotométrica de las mismas, con el objetivo de dar a todas ellas un aspecto global similar. Finalmente, y con el objetivo de eliminar, en el mayor grado posible, cualquier artefacto fruto de errores en el registro, las diferencias fotométricas no solventadas o a la manifestación del fenómeno del paralaje, se realizará una etapa de fusión. En dicha etapa se pretende otorgar a la imagen de mosaico resultante una apariencia lo más continua y convincente posible.

#### 3.1. Registro geométrico

Dada la secuencia de imágenes con la que se ha trabajado, en primer lugar se ha procedido al registro entre parejas de

imágenes consecutivas, con el objetivo de obtener una primera aproximación de la topología del mosaico. Para ello, se ha aplicando un método basado en la búsqueda de puntos característicos y sus correspondencias, y concretamente se han utilizado el detector de puntos de Harris (Harris and Stephens (1988)) y el descriptor SURF (Bay et al. (2006)). En cada una de las imágenes se han buscado una serie de puntos característicos que, una vez identificados en la imagen posterior de la secuencia, han permitido la estimación de una transformación geométrica entre las mismas, representada por una matriz denominada homografía (Hartley and Zisserman (2003)). Durante la detección de correspondencias es posible que algunas de ellas no sean correctas, ya sea por presencia de ruido en las imágenes, o porque no obedecen al movimiento dominante entre las mismas, por pertenecer a objetos en movimiento. Dichas correspondencias, conocidas como *outliers*, han sido eliminadas mediante el algoritmo de RANSAC propuesto por Fischler and Bolles (1981).

Una vez dispuestas todas las imágenes en un mismo marco, y por lo tanto acotadas las vecindades de búsqueda, ha sido posible localizar correspondencias entre imágenes no consecutivas, necesarias para la aplicación de un algoritmo de alineamiento global (Ferrer et al. (2007); Elibol et al. (2008)). El alineamiento global tiene en cuenta las correspondencias entre imágenes consecutivas, aquellas entre imágenes no consecutivas, y las coordenadas fiduciales de que se pueda disponer, que en el caso que nos ocupa han sido dos. Los puntos fiduciales permiten situar únicamente en el globo el mosaico generado, y son interpretados como restricciones geométricas, durante el proceso de ajuste de las posiciones de todas las imágenes. El alineamiento global se traduce en la minimización de los residuos de la función de coste que se presenta a continuación:

$$\operatorname{argmin} \left( \tau \cdot \sum_{\substack{1 \leq l \leq m \\ 1 \leq k \leq n}} (\|{}^l r_k^l\|^2 + \|{}^j r_k^l\|^2) + \nu \cdot \sum_{k=1}^o \|{}^p r_k\|^2 \right) \quad (1)$$

donde  $\tau$  y  $\nu$  son pesos de ponderación,  $m$  es el número de correspondencias existentes entre una pareja de imágenes,  $n$  es el número de parejas de imágenes con superposición y  ${}^l r_k^l$ ,  ${}^j r_k^l$  y  ${}^p r_k$  se definen como:

$${}^l r_k^l = {}^l x_k^l - {}^l H \cdot {}^j x_k^l \quad (2)$$

$${}^j r_k^l = {}^j x_k^l - {}^j H \cdot {}^l x_k^l \quad (3)$$

$${}^p r_k = {}^l x_k^l - {}^m \tilde{x}_0^l \quad (4)$$

El primer término de la Ecuación 1 representa el error de reproyección de las correspondencias, es decir, aquél que mide la distancia entre el resultado de aplicar la transformación entre imágenes sobre los puntos de interés y la posición real de las correspondencias en ambos sentidos de la transformación. Así, en la Ecuación 2,  ${}^l x_k^l$  es un punto detectado en la imagen  $i$  mientras que  ${}^j x_k^l$  es su correspondencia en la imagen  $j$ . Ambos puntos están expresados en coordenadas homogéneas, y  ${}^l H$  representa la transformación plana entre ambas imágenes, mientras que los papeles se intercambian en la Ecuación 3. El segundo término de la Ecuación 1 hace referencia a los puntos fiduciales conocidos, que nos permiten georeferenciar todo el mosaico, y mide

la desviación en la posición de un determinado punto fiducial  ${}^l x_k$ , capturado en la imagen  $i$ , respecto de la posición real  ${}^m \tilde{x}_0^l$  en el plano del mosaico.

La importancia del alineamiento global emerge en las situaciones de cierre de lazos, dado que permiten mejorar significativamente la trayectoria de la cámara cuando se vuelven a visitar áreas ya mapeadas previamente. En ausencia de situaciones de cierre de lazos, y teniendo en cuenta que las secuencias pueden estar constituidas por miles de imágenes, la deriva acumulada por las transformaciones entre parejas conduce a la obtención de mosaicos ostensiblemente inconsistentes, con desarrollos como el de la Figura 1.

### 3.2. Procesado fotométrico

#### **Estimación de la calidad de las imágenes**

En la práctica no existe un criterio único y objetivo que permita identificar la imagen con más calidad visual, dada una cierta secuencia de las mismas, puesto que el concepto de "calidad" implica diferentes aspectos cognitivos. Sin embargo, lo que sí que es cierto es que dos imágenes adquiridas con la misma distancia focal, aquella que haya sido adquirida a menor distancia de la escena presentará un nivel de detalle más elevado que la que haya sido adquirida desde más lejos. Según lo dicho, la distancia a la escena podría ser utilizada como medida para indexar la calidad de las imágenes adquiridas.

Un método más robusto de indexado de las imágenes consistiría en utilizar la medida de *Total Variation* Osher and Rudin (1990); Dibos and Koepfler (1999) como evaluador de calidad. La medida de *Total Variation* se denota como  $T_V$  y su valor se define de la siguiente manera:

$$T_V = \frac{1}{W \cdot H} \sum_{x=1}^{W-1} \sum_{y=1}^{H-1} \|g(x, y)\| \quad (5)$$

donde  $W$  y  $H$  son las dimensiones horizontal y vertical de la imagen y  $g(x, y)$  corresponde al vector de gradiente del píxel  $(x, y)$ .  $T_V$  evalúa la magnitud media de los valores de gradiente, otorgando una mayor puntuación a aquellas imágenes con las transiciones más definidas entre píxeles colindantes, y por lo tanto, más nítidas. Por el contrario, las imágenes borrosas reciben una puntuación baja, ya que las transiciones entre píxeles colindantes son más difusas.

#### **Extensión global del contraste**

Una vez seleccionada una imagen como referencia, en función de la distancia de adquisición de la cámara respecto de la escena o de su valor de  $T_V$ , es necesario evaluar el contraste de la misma. El valor  $T_V$  es utilizado para ello, así como para calcular la ganancia que será aplicada para obtener contraste global (o amplificación de gamma) en el resto de imágenes. La ganancia debe encontrarse sobre un umbral  $T_S$ , con el objetivo de evitar una amplificación excesiva de las áreas de pobre contraste, e.g., regiones monótonas sin textura.  $T_S$  dependerá del ratio señal-ruido de la imagen, que puede variar notablemente en función de las condiciones meteorológicas durante la adquisición. Un valor experimental de  $T_S = 5$  ha sido utilizado du-

rante los experimentos. El factor de extensión  $\frac{T_{V_{referencia}}}{T_{V(k)}}$  se aplica sobre los componentes  $x$  e  $y$  de los gradientes de la  $k$ -ésima imagen.

En el caso en que las diferencias de iluminación no sean únicamente fruto de diferentes tiempos de exposición, sino que dependan de otros factores, como la temperatura del color (diferente entre el amanecer y el atardecer, por ejemplo), se hace además necesario efectuar una clonación de estilo entre las imágenes capturadas. Así, es posible utilizar un método como el de Reinhard et al. (2001), que permite transferir el estilo de color de una de las imágenes al resto de la secuencia, con tal de obtener el aspecto más uniforme posible del mosaico final.

### 3.3. Fusión de las imágenes

El fenómeno del paralaje puede afectar tanto al proceso de registro geométrico como al de fusión de las imágenes. La existencia de paralaje, en aquellas secuencias adquiridas por una cámara que haya realizado un movimiento de traslación, puede dificultar el correcto registro geométrico entre imágenes adyacentes. Así mismo, incluso cuando el registro obtenido sea muy preciso, existe la posibilidad de que aparezcan dobles contornos, en las regiones comunes superpuestas, debido a los diferentes puntos de vista desde los que se han capturado los objetos o estructuras de la escena. La ausencia de información de calibración de los parámetros intrínsecos de la cámara, que permita corregir distorsiones debidas a la lente, también puede dar lugar a dicho artefacto durante la fusión.

La solución para evitar los artefactos de doble contorno consiste en utilizar información de una única imagen para cada uno de los píxeles del mosaico construido, siempre que esto sea posible. Con el objetivo de reducir, en la mayor medida posible, los efectos en el área de superposición, obteniendo a su vez una transición suave entre imágenes, la fusión se realiza únicamente en una región poco extensa (de unos pocos píxeles) alrededor de una costura óptima previamente calculada. Consecuentemente, la proporción del mosaico en la que la información proviene de más de una imagen es ínfima, y por lo tanto la posibilidad de aparición de artefactos en las áreas de superposición se ve muy reducida.

#### Primer y segundo mapas de proximidad

El método de suavizado propuesto requiere la determinación del primer y segundo mapas de proximidad a nivel de pixel. El primer mapa de proximidad contiene, para cada coordenada de píxel, el índice de la imagen cuyo centro es más cercano. El segundo mapa de proximidad, como su nombre indica, contiene el índice de la imagen cuyo centro es el segundo más cercano a dicho píxel. De forma similar a Agarwala et al. (2004), se aplica un algoritmo de corte de grafo sobre la superposición entre ambos mapas, que permite calcular los bordes de las imágenes y las regiones de suavizado establecidas a su alrededor. Para cada píxel del borde y de la región de suavizado se seleccionan dos índices. Consecuentemente, cada pixel ubicado fuera de dicha área (es decir, los correspondientes a la mayor parte del mosaico) es asociado a una única imagen.

La distancia Euclídea entre un pixel  $I^M(x, y)$ , expresado en el sistema de coordenadas del mosaico, y el centro de la  $n$ -ésima

imagen  $I^n(x, y)$ , se pondera por un factor  $w_n(s)$ , tal y como se muestra en la Ecuación 6:

$$d_M^m(x, y) = w_n(s) \cdot \sqrt{(x_M - x_n)^2 + (y_M - y_n)^2} \quad (6)$$

donde el escalar  $w_n(s)$  es un cociente entre el tamaño de la imagen  $n$ -ésima y la imagen que presenta un tamaño más pequeño, una vez registrada. Por razones de eficiencia temporal, el cociente no se calcula basándose en el área de las imágenes, una vez transformadas, sino en la longitud de sus diámetros, con el objetivo de obtener únicamente una aproximación, de la forma que se muestra en la Ecuación 7:

$$w_n(s) = s_{min}/s_n \quad (7)$$

donde  $s_{min}$  es el diámetro de la imagen más pequeña de un conjunto dado, y  $s_n$  es el diámetro de la  $n$ -ésima imagen.

La ponderación presentada prioriza los píxeles de las imágenes adquiridas a menor altitud y que, por lo tanto, se puede considerar que presentan un mayor nivel de detalle. Así, en aquellos casos en los que una imagen adquirida a una menor altura se encuentre superpuesta con otra adquirida a mayor altura, la primera imagen será la que aportará la mayor parte de los píxeles para la región común.

#### Regiones de intersección

La superposición entre el primer y el segundo mapas de proximidad determina las regiones donde el corte de grafo a nivel de píxel puede ser efectuado. De tal modo, para cada región de superposición, la textura proveniente de las dos imágenes con mayor calidad puede ser recuperada. El corte de grafo puede ser, a su vez, calculado con tal de obtener el límite óptimo entre las mismas, que determinará la contribución de ambas al mosaico final. Se define  $ROI_{i,j}$  como la región de intersección entre las imágenes  $i$  y  $j$ , donde  $i$  es la imagen más cercana,  $j$  es la segunda imagen más cercana y  $R$  denota la región del mosaico donde coinciden simultáneamente  $i$  y  $j$ . Es decir,  $ROI_{i,j} = R_{i,j} \cup R_{j,i}$ .

#### Corte de grafo a nivel de píxel

La estrategia de fusión propuesta utiliza un algoritmo de búsqueda de borde óptimo, para calcular los mejores límites de las imágenes en las regiones de superposición. Se realiza un corte de grafo a nivel de píxel en las regiones de intersección, determinadas por el primer y segundo mapas de proximidad. El algoritmo busca un camino que minimice el coste de la transición, de un lado al otro del borde, para cada pareja de píxeles colindantes. La función tiene tres términos ponderados, que controlan el comportamiento del corte:

$$C = \mu_1 \cdot f(I_1, I_2) + \mu_2 \cdot s(g_1, g_2) + \mu_3 \cdot L \quad (8)$$

El primer término  $\mu_1 \cdot f(I_1, I_2)$  mide las diferencias de intensidad entre los píxeles superpuestos. El segundo término  $\mu_2 \cdot s(g_1, g_2)$  mide los gradientes a lo largo del borde  $B$ . Finalmente, el tercer término  $\mu_3 \cdot L$  mide la longitud  $L$  del camino. Como aportación en el presente trabajo, nuestro método utiliza el término de gradiente  $s(g_1, g_2)$ , en la función de coste que controla el corte, que mide el coste de la transición de un lado al

otro del borde, a medida que el camino óptimo se va calculando. El término de gradiente es combinado con el término de intensidad  $f(I_1, I_2)$ , y se diferencia de otros métodos de la bibliografía como el de Mills and Dudek (2009), en el que el corte se realiza, o bien en el dominio de las diferencias de intensidad, o bien en el dominio de las diferencias en el gradiente, de manera excluyente la una de la otra. El término de gradiente permite tratar con imágenes adquiridas con diferentes tiempos de exposición, en las regiones de superposición. En dicho caso, un corte de grafo basado únicamente en los valores de intensidad consideraría que las diferencias entre píxeles vecinos son grandes, incluso si el registro es preciso, y consecuentemente el corte calculado no sería el más adecuado. En cambio, si se utiliza la diferencia entre los vectores de gradiente, el camino óptimo podrá ser encontrado, independientemente de los diferentes tiempos de exposición de las imágenes. Además, el hecho de trabajar en el dominio del gradiente permite compensar las diferencias de exposición, en el momento de recuperar la luminancia.

Mediante esta estrategia (ver Figura 6) se minimizan los efectos del paralejo y las imprecisiones en el registro geométrico, dado que el corte de grafo tiende a situar el borde en regiones sin textura, donde las diferencias morfológicas son bajas. Por la misma razón, los cortes sobre objetos en movimiento tienden a ser evitados, beneficiando dicho comportamiento la consistencia visual del resultado fusionado.

Realizar un corte de grafo, especialmente a nivel de píxel, es habitualmente una operación costosa, en términos computacionales, cuando la medida de la región a procesar es significativamente grande. Sin embargo, las regiones en las cuales realizamos el corte de grafo, determinadas por la intersección entre el primer y segundo mapas de proximidad, son raramente extensas. Además, dicho proceso puede ser paralelizado, usando las ventajas de los recientes procesadores de múltiples núcleos, con tal de reducir el tiempo de ejecución de uno de los más importantes cuellos de botella de la secuencia de proceso.

#### Fusión en el dominio del gradiente

Una vez el borde óptimo ha sido localizado, será necesario obtener una transición suave entre las regiones adyacentes. Incluso en el caso de secuencias en las que las imágenes han sido procesadas, para reducir los efectos de posibles diferencias de iluminación y exposición, y en las cuales se ha realizado una ecualización de histograma, el corte de grafo puede conducir a una imagen con bordes visibles. Por lo tanto, el suavizado de la transición entre las diferentes regiones de la imagen es fundamental. La fusión de imágenes alrededor de los bordes calculados debe ser realizada en un área limitada, lo suficientemente ancha como para asegurar una transición suave y no perceptible, pero también lo suficientemente estrecha como para reducir la posible visibilidad de regiones borrosas o dobles contornos.

El método propuesto consiste en la aplicación de un promedio ponderado alrededor de los bordes en el dominio del gradiente, tal y como se propone en la Ecuación 9:

$$\begin{aligned}\hat{g}_x(x, y) &= \mu g_x^1(x, y) + (1 - \mu) \cdot g_x^2(x, y) \\ \hat{g}_y(x, y) &= \mu g_y^1(x, y) + (1 - \mu) \cdot g_y^2(x, y)\end{aligned}\quad (9)$$

donde  $g_x^1$ ,  $g_y^1$ ,  $g_x^2$  y  $g_y^2$  son las matrices de gradiente en las direcciones  $x$  e  $y$ , correspondientes a las dos imágenes involucradas,  $\hat{g}_x$  y  $\hat{g}_y$  son las matrices de gradiente en las direcciones  $x$  e  $y$ , después de la fusión, y  $\mu$  es la función de transición de suavizado, concretamente una función de Hermite de tercer orden. La ventaja de aplicar un promedio ponderado en el dominio del gradiente es la compensación automática de las diferentes exposiciones entre imágenes vecinas, cuando se integra la luminancia desde los gradientes, como etapa final.

#### Recuperación de la luminancia desde el gradiente

Después de procesar independientemente cada una de las áreas correspondientes a las regiones de superposición alrededor de los bordes, los fragmentos resultantes necesitan ser unificados en una única imagen, de mayores dimensiones. Cada fragmento procesado deberá ser actualizado en la imagen del mosaico final, mientras que la información perteneciente a regiones sin superposición debe ser recuperada de las imágenes originales pertinentes. Dicha forma de proceder permite conservar el mayor nivel de detalle posible en aquellas regiones en las que la fusión, susceptible de producir efectos visuales no deseables, no sea necesaria.

Una vez el mosaico final en el dominio del gradiente ha sido compuesto, después de la fusión alrededor de los bordes, se obtiene un campo de gradiente no integrable o inconsistente. Para recuperar los valores de luminancia de los campos de gradiente  $G_x$  y  $G_y$ , es necesario resolver la ecuación de Poisson  $\nabla^2 I = \tilde{G}$  (Fattal et al. (2002)).  $\nabla^2$  es el operador Laplaciano  $\nabla^2 = \frac{\delta^2 I}{\delta x^2} + \frac{\delta^2 I}{\delta y^2}$  y  $\tilde{G}$  es la divergencia del campo de vectores  $G$ , definida como  $\tilde{G} = \frac{\delta G_x}{\delta x} + \frac{\delta G_y}{\delta y}$ .

La resolución del sistema de ecuaciones diferenciales requiere la especificación de condiciones de contorno. Se utilizan condiciones de contorno de tipo Neumann, definidas como  $\nabla I \cdot \mathbf{n} = 0$ , i.e., la derivada en la dirección normal  $\mathbf{n}$  es cero. El sistema es resuelto utilizando un *multigrid Poisson solver* (Kazhdan and Hoppe (2008)). La solución proporcionada por el algoritmo se encuentra definida respecto a un término aditivo, en relación a los valores de intensidad recuperados. Consecuentemente, un algoritmo de mapeo, como *Minimum Information Loss* (Neumann et al. (1998)), debe ser aplicado para obtener dicho factor. El objetivo principal del algoritmo de mapeo es el de manipular apropiadamente el rango dinámico de la imagen calculada, con tal de permitirle encajar de la mejor manera posible en el rango limitado de un dispositivo de presentación, manteniendo a su vez la mayor parte de la información de detalle.

#### 4. Resultados y discusión

La Figura 2 muestra algunas imágenes de 39 Mpíxeles ( $6216 \times 5412$ ) utilizadas para la construcción del mosaico (ver Figura 5). Cabe destacar la diferencia de aspecto de la superficie del agua, que presenta a su vez iluminación no uniforme, debido a la posición del sol respecto a la cámara, en el momento de la adquisición, durante toda la secuencia (ver Figura 4).

El mosaico ha sido generado, durante el registro geométrico, en una primera iteración, utilizando SURF, tanto para la detección de puntos característicos (permitiendo un máximo de 3000 por imagen durante la búsqueda) como para su descripción, mientras que en una segunda ejecución los resultados de la detección han sido refinados mediante Harris. Los puntos característicos han sido distribuidos uniformemente sobre todo el área de las imágenes, ya que el porcentaje de superposición entre las mismas es elevado. Han sido utilizadas 23794 correspondencias para la ejecución del alineamiento global, modelado sobre transformaciones con seis grados de libertad, que ha dado lugar a un error de reproyección medio de un 0,083 % respecto del tamaño total de la imagen. En dicho alineamiento global se han utilizado los valores  $\tau = 0,5$  y  $\nu = 0,5$ , con el fin de otorgar la misma importancia tanto a los puntos característicos encontrados en las imágenes como a los puntos de control fiduciales marcados manualmente. En cuanto a los parámetros que controlan el corte de grafo a nivel de pixel, los valores utilizados han sido  $\mu_1 = 0,4$ ,  $\mu_2 = 0,4$  y  $\mu_3 = 0,2$ . Los valores han sido determinados experimentalmente, basándose en la idea que las imágenes a registrar tienen, por lo general, una apariencia similar y, por lo tanto, los pesos relativos a las diferencias fotométricas y de gradientes deben ser similares. Así mismo, el peso asociado a la longitud del corte tiene un peso inferior al del resto, ya que en el caso que nos ocupa la longitud de dicho corte, dado el buen registro de las imágenes, no es especialmente relevante. En aquellos casos en los que las diferencias en los tiempos de exposición entre imágenes sean más acusadas, un mayor valor del factor que controla el gradiente permitirá un mejor corte, dado que el gradiente no es sensible a la exposición. Por lo que se refiere al tercer parámetro, un valor excesivamente bajo podría resultar en un corte demasiado largo y complejo, que podría repercutir en la calidad del suavizado posterior, por lo que en ningún caso es aconsejable prescindir de dicho factor. Los resultados son convincentes, teniendo en cuenta la gran dimensión de las imágenes y el hecho de que no se ha podido realizar la calibración de la cámara (Tsai (1986); Faugeras et al. (1992)) y, por lo tanto, no han podido ser corregidas las distorsiones provocadas por la lente.

En lo referente a la calidad de la fusión, se puede observar que las diferencias de exposición entre imágenes han sido compensadas, y que las uniones entre imágenes tampoco son perceptibles. En dichas uniones no hay dobles contornos o regiones borrosas gracias al uso de un corte de grafo para la selección óptima de los límites y la aplicación del suavizado en el dominio del gradiente, en una región relativamente estrecha alrededor de los mismos. El uso de técnicas de suavizado en el dominio del gradiente ha permitido que se hayan resuelto con solvencia no sólo los diferentes tiempos de exposición, sino también los aspectos dispares entre imágenes colindantes, como las de la Figura 4. Algoritmos como Burt and Adelson (1983) no son capaces de trabajar sobre escenarios con paralaje ni diferentes condiciones de iluminación (ver Figura 6), mientras que el método propuesto funciona satisfactoriamente en dichos escenarios. El problema de más difícil solución es la presencia de agua encrespada en el lago, por efecto del viento. El aspecto del agua en las imágenes afectadas dista significativamente del

de las imágenes no afectadas. No obstante, la transición obtenida por el método propuesto es suave y permite que el mosaico conserve una apariencia continua y consistente.



Figura 4: Detalle de una región central del mosaico del lago de Banyoles antes (a) y después (b) de la aplicación de la técnica de fusión. En el área del detalle contribuyen un total de seis imágenes, con evidentes diferencias de exposición. La imagen resultante es totalmente continua, sin visibilidad de costuras, y la exposición es constante en todo el área.

## 5. Conclusiones y trabajo futuro

La construcción de ortofotomapas, a partir de imágenes adquiridas desde un vehículo aéreo, es una tarea con una presencia razonable en la literatura. Sin embargo, la obtención de una imagen con apariencia de alta calidad no ha sido siempre un objetivo prioritario, como sí lo ha sido en el contexto de la generación de panoramas de imágenes. El objetivo del trabajo ha sido presentar toda una secuencia de proceso capaz de obtener, de forma automática, ortofotomapas de alta calidad, sin artefactos geométricos ni fotométricos entre las imágenes fusionadas.

El sistema propuesto da lugar a ortofotomapas georeferenciados, y el registro geométrico entre las imágenes es refinado mediante la aplicación de un algoritmo de alineamiento global, que a su vez tiene en consideración puntos fiduciales, manualmente marcados en algunas imágenes. Dicho algoritmo permite garantizar la consistencia del mosaico generado, así como evitar la degeneración geométrica sufrida por los métodos de registro secuenciales. En lo que se refiere al aspecto fotométrico, la utilización de un algoritmo de corte de grafo, a nivel de pixel, permite obtener emplazamientos precisos para las costuras que minimizan las diferencias fotométricas de la transición, así como obtener transiciones suaves en escenas con paralaje moderado. Además, dichas transiciones son suavizadas posteriormente mediante un método de fusión en el dominio del gradiente que, además de conseguir una transición no perceptible, permite también la compensación de las posibles diferencias de iluminación o tiempos de exposición entre imágenes. Se ha propuesto un mecanismo de priorización de las imágenes que otorga mayor peso a aquellos píxeles adquiridos a menor distancia del suelo y, por consiguiente, corresponden a una imagen con mayor resolución y afectada en menor medida por los fenómenos atmosféricos. Dicha ponderación es utilizada tanto durante el cálculo de las costuras óptimas como durante la selección de la contribución de las imágenes al ortofotomap. Se ha demostrado la aplicación del algoritmo propuesto a partir de un vuelo aéreo sobre el lago de Banyoles, dando lugar a un mosaico en color de alta resolución, de apariencia consistente y sin artefactos geométricos ni fotométricos, que resultará de utilidad

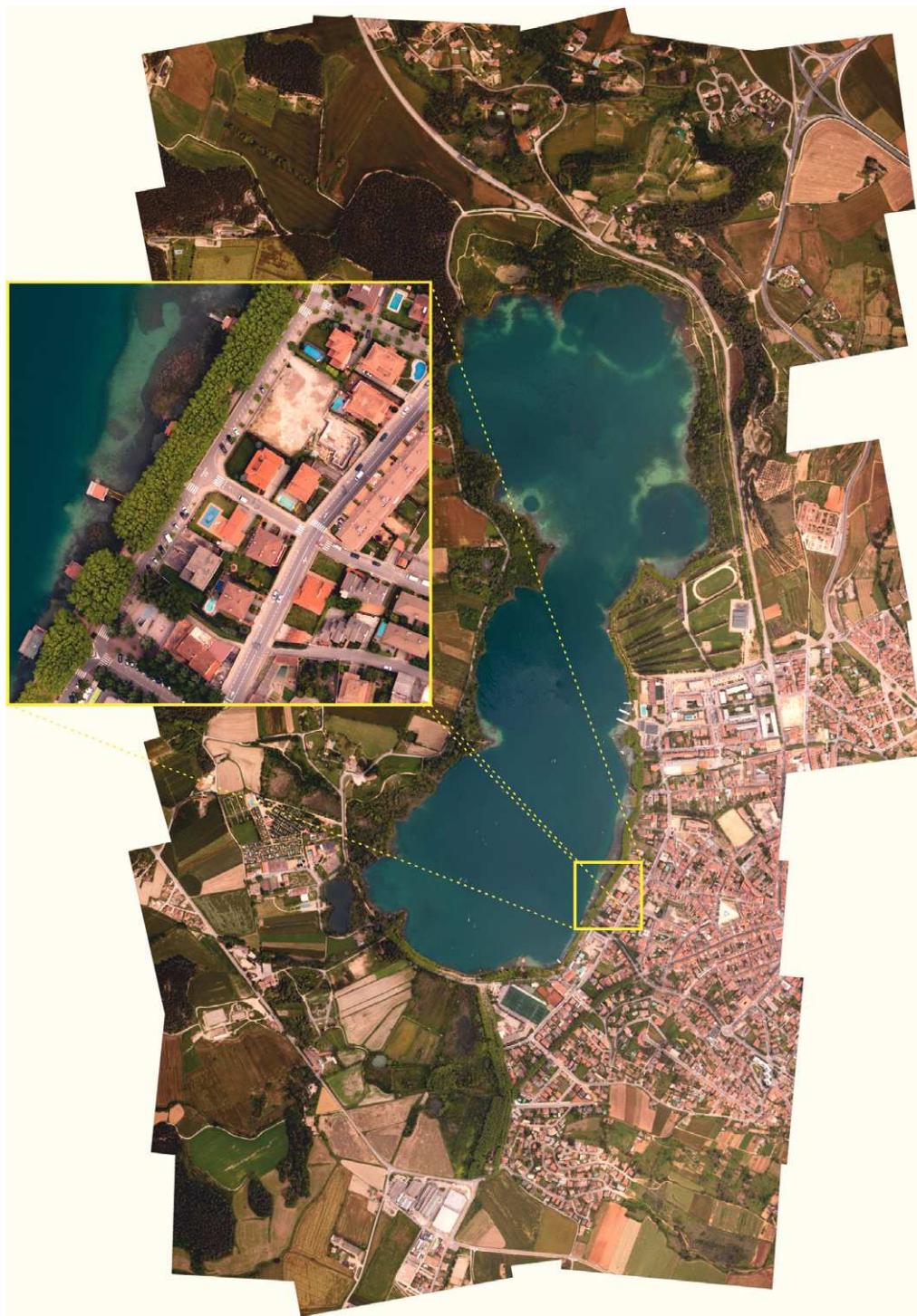


Figura 5: Aspecto del mosaico de Banyoles construido, una vez ejecutadas las etapas de registro geométrico, con alineamiento global, y de fusión. Las transiciones entre imágenes resultan imperceptibles en la práctica. Únicamente se vislumbra la diferencia de textura de la superficie del lago, en una región correspondiente a una imagen en la que el agua se encontraba ligeramente crispada. Las dimensiones del mosaico son  $23740 \times 39076$  píxeles y las de la selección  $2000 \times 2000$  píxeles.



Figura 6: Comparación entre los resultados obtenidos por el algoritmo Burt and Adelson (1983) y el método propuesto. Las dos imágenes a la izquierda fueron adquiridas con una diferencia de aproximadamente un año (Girona, 2010 y 2011), y su apariencia es dispar. El leve paralelo de la escena provoca que el método de Burt and Adelson (primer mosaico) muestre dobles contornos alrededor de los edificios de la zona central. El método propuesto (segundo mosaico) no sólo no es sensible a dicho efecto, sino que además permite compensar las diferencias en la iluminación del ambiente (Fuente: Instituto Cartográfico de Catalunya).

para el estudio geológico de los márgenes del mismo. Aunque el presente trabajo se ha centrado en el procesamiento de una secuencia de imágenes aéreas, el método propuesto es genérico y presenta múltiples aplicaciones, como la construcción de panoramas o la generación de mosaicos submarinos. Los mecanismos adaptativos de corrección de las imágenes permiten que el algoritmo obtenga buenos resultados independientemente del origen de las imágenes adquiridas.

Como extensión al trabajo desarrollado, resulta interesante tratar el problema de la unificación de superficie del agua, en casos en los que su apariencia varía considerablemente entre imágenes adyacentes, debido a fenómenos atmosféricos como el viento. Dicho tratamiento requeriría, muy probablemente, la utilización de métodos de segmentación para el proceso específico de dicho elemento.

### English Summary

#### Automatic construction of ortophotomaps: a photometric approach

#### Abstract

Mosaicing allows to obtain a high-resolution representation of a given scene. Off-the-shelf still cameras including built-in software to build photo-mosaics and online applications such as Google Maps allowing to visualize maps resulting from a photo-mosaic are common nowadays. In most cases panoramic mosaics are generated from images acquired by means of a camera undergoing uniquely a rotation movement. When the acquisition conditions change, and the camera also performs a translation movement, the parallax phenomenon appears. If parallax exists, the seamless combination of the images is even more challenging. Additionally, adjacent photographs may present

differences in appearance due to some automatic camera mechanisms, such as the automatic exposure. In this work a full processing pipeline intended to automatically build seamless mosaics with continuous and consistent appearance is described. Images are joined using methods which guarantee geometrical consistency, and fused using blending techniques, to achieve a non-visible transition between images. The described pipeline is applied on a high-resolution image sequence in order to evaluate its application in the context of aerial images of large dimensions.

#### Keywords:

Image processing, image enhancement, image matching, image registration, gradient methods

#### Agradecimientos

El trabajo ha sido parcialmente financiado por el MICINN bajo el proyecto CTM2010-15216. László Neumann ha sido financiado por ICREA de la Generalitat de Catalunya. Las imágenes de las Figuras 2, 4 y 5 son gentileza del Consorcio de l'Estany de Banyoles.

#### Referencias

- Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S., Colburn, A., Curless, B., Salesin, D., Cohen, M., August 2004. Interactive digital photomontage. In: Proc. SIGGRAPH04.
- Arévalo, V., Gonzalez, J., 2008. An Experimental Evaluation of Non-Rigid Registration Techniques on QuickBird Satellite Imagery. International Journal of Remote Sensing 29 (2), 513–527.
- Bay, H., Tuytelaars, T., Van Gool, L. J., 2006. SURF: Speeded Up Robust Features. In: European Conference on Computer Vision. pp. 404–417.
- Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C., July 2000. Image inpainting. In: Proceedings of SIGGRAPH. New Orleans, USA. pp. 417–424.
- Botterill, T., Mills, S., Green, R., 2009. New conditional sampling strategies for speeded-up RANSAC. In: BMVC.

- Botterill, T., Mills, S., Green, R., November 2010. Real-time Aerial Image Mosaicing. In: Proceedings of Image and Vision Computing New Zealand. Queenstown, NZ, pp. 1–6.
- Burt, P., Adelson, E., 1983. A multiresolution spline with application to image mosaics. ACM Trans. Graph. 2 (4), 217–236.
- Dibos, F., Koepfler, G., December 1999. Global total variation minimization. SIAM J. Numer. Anal. 37, 646–664.
- Dijkstra, E. W., 1959. A note on two problems in connexion with graphs. Numerische Mathematik 1, 269–271.
- Ding, M., Lyngbaek, K., Zakhor, A., June 2008. Automatic registration of aerial imagery with untextured 3d lidar models. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. pp. 1–8.
- Elibol, A., Garcia, R., Delaunoy, O., Gracias, N., 2008. A new global alignment method for feature based image mosaicing. In: ISVC '08: Proceedings of the 4th International Symposium on Advances in Visual Computing, Part II. Springer-Verlag, Berlin, Heidelberg, pp. 257–266.
- Elibol, A., Gracias, N., Garcia, R., Sep. 2010. Augmented state-extended kalman filter combined framework for topology estimation in large-area underwater mapping. J. Field Robot. 27 (5), 656–674.  
URL: <http://dx.doi.org/10.1002/rob.v27:5>  
DOI: 10.1002/rob.v27:5
- Escartín, J., García, R., Delaunoy, O., Ferrer, J., Gracias, N., Elibol, A., Cufí, X., Neumann, L., Fornari, D. J., Humphris, S. E., Renard, J., December 2008. Globally aligned photomosaic of the lucky strike hydrothermal vent field (mid-atlantic ridge, 37°deg18.5'N): Release of georeferenced data, mosaic construction, and viewing software. Geochemistry, Geophysics, Geosystems 9, Q12009.
- Fattal, R., Lischinski, D., Werman, M., 2002. Gradient domain high dynamic range compression. ACM Trans. Graph. 21 (3), 249–256.
- Faugeras, O., Luong, Q.-T., Maybank, S., May 1992. Camera self calibration : theory and experiments. In: Proc. of the 2nd. European Conference on Computer Vision. Santa Margherita, Italy.
- Ferrer, J., Elibol, A., Delaunoy, O., Gracias, N., Garcia, R., 2007. Large-area photo-mosaics using global alignment and navigation data. In: MTS/IEEE OCEANS.
- Fischler, M., Bolles, R., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 6 (24), 381–395.
- Girosi, F., Verri, A., Torre, V., 1989. Constraints for the computation of optical flow. In: Proc. of the IEEE Workshop on Visual Motion. pp. 116–124.
- González, J., Ambrosio, G., Arévalo, V., Nov. 2001. Automatic Urban Change Detection from the IRS-1D PAN. In: IEEE-ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas. Rome, Italy, pp. 320–323.
- Harris, C., Stephens, M., August 1988. A combined corner and edge detector. In: Proceedings Alvey Conference. Manchester, UK, pp. 189–192.
- Hartley, R., Zisserman, A., 2003. Multiple view geometry in computer vision, 2nd edition. Cambridge University Press.
- Horn, B., Shunck, B., 1981. Determining optical flow. Artificial Intelligence 17, 185–203.
- Kanatani, K., Ohta, N., September 1999. Accuracy bounds and optimal computation of homography for image mosaicing applications. In: Proc. of the Seventh International Conference on Computer Vision. Vol. 1. IEEE, pp. 73–78.
- Kazhdan, M., Hoppe, H., 2008. Streaming multigrid for gradient-domain operations on large images. ACM Trans. Graph. 27 (3), 1–10.
- Kumar, R., Anandan, P., Irani, M., Bergen, J., Hanna, K., June 1995. Representation of scenes from collections of images. In: Proc. of the IEEE Computer Society Workshop on Visual Scenes Representation. Cambridge, MA.
- Kumar, R., Sawhney, H., Asmuth, J., Pope, A., Hsu, S., August 1998. Registration of video to geo-referenced imagery. In: Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on. Vol. 2. pp. 1393–1400 vol.2.
- Lear, A., sep/oct 1997. Digital orthophotography: mapping with pictures. Computer Graphics and Applications, IEEE 17 (5), 12–14.
- Levin, A., Zomet, A., Peleg, S., Weiss, Y., May 2004. Seamless image stitching in the gradient domain. In: Proc. of the European Conference on Computer Vision (ECCV04). Prague, Czech Republic.
- Lindeberg, T., 1993. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. International Journal of Computer Vision 11, 283–318, 10.1007/BF01469346.
- Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60 (2), 91–110.
- Lowe, D. G., 1999. Object recognition from local scale-invariant features. In: ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2. IEEE Computer Society, Washington, DC, USA, p. 1150.
- Milgram, D., 1975. Computer methods for creating photomosaics. IEEE Transactions on Computers 24 (11), 1113–1119.
- Milgram, D., November 1977. Adaptive techniques for photomosaicking. IEEE Transactions on Computers C-26 (11), 1175–1180.
- Mills, A., Dudek, G., 2009. Image stitching with dynamic elements. Image and Vision Computing 27 (10), 1593–1602.
- Morse, B. S., Engh, C. H., Goodrich, M. A., 2010. Uav video coverage quality maps and prioritized indexing for wilderness search and rescue. In: Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction. HRI '10. ACM, New York, NY, USA, pp. 227–234.
- Naydenova, V., Jelev, G., June 2009. Forest dynamics study using aerial photos and satellite images with very high spatial resolution. In: Recent Advances in Space Technologies, 2009. RAST '09. 4th International Conference on. pp. 344–348.
- Neumann, L., Matkovic, K., Purgathofer, W., 1998. Automatic exposure in computer graphics based on the minimum information loss principle. Computer Graphics International Conference 0, 666–667.
- Nicosevici, T., Garcia, R., October 2009. On-line visual vocabularies for robot navigation and mapping. In: Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on. pp. 205–212.
- Nicosevici, T., Gracias, N., Negahdaripour, S., Garcia, R., 2009. Efficient three-dimensional scene modeling and mosaicing. J. Field Robot. 26 (10), 759–788.
- Osher, S., Rudin, L., Aug. 1990. Feature-oriented image enhancement using shock filters. SIAM J. Numer. Anal. 27 (4), 919–940.
- Peleg, S., 1981. Elimination of seams from photomosaics. In: Pattern Recognition and Image Processing. Vol. 16. pp. 90–94.
- Pérez, P., Gangnet, M., Blake, A., 2003. Poisson image editing. ACM Transactions on Graphics (SIGGRAPH'03) 22 (3), 313–318.
- Reinhard, E., Adhikmin, M., Gooch, B., Shirley, P., sep/oct. 2001. Color transfer between images. Computer Graphics and Applications, IEEE 21 (5), 34–41.
- Solberg, R., Solberg, A., Koren, H., Aas, K., May 1996. The suitability of future high-resolution satellite imagery for forest inventory. In: Geoscience and Remote Sensing Symposium, 1996. IGARSS '96. 'Remote Sensing for a Sustainable Future', International. Vol. 4. pp. 2307–2309 vol.4.
- Tegolo, D., Valenti, C., 2001. A naive approach to compose aerial images in a mosaic fashion. Image Analysis and Processing, International Conference on on 0, 0512.
- Tsai, R., 1986. An efficient and accurate camera calibration technique for 3D machine vision. In: IEEE International Conference on Computer Vision and Pattern Recognition - CVPR86. Miami Beach, Florida.
- Wehn, H., Goldstein, N., Ameri, B., Moshkovitz, A., Zwick, H., june 2002. Frequent-image-frames enhanced digital ortho-rectified mapping (fifedom) airborne mapping system. In: Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International. Vol. 2. pp. 1296–1298 vol.2.
- Wen, H., Zhou, J., dec. 2008. An improved algorithm for image mosaic. In: Information Science and Engineering, 2008. ISISE '08. International Symposium on. Vol. 1. pp. 497–500.
- Zebedin, L., Klaus, A., Gruber-Geymayer, B., Karner, K., 2006. Towards 3d map generation from digital aerial images. ISPRS Journal of Photogrammetry and Remote Sensing 60 (6), 413–427, digital Aerial Cameras.
- Zhao, W., 2006. Flexible image blending for image mosaicing with reduced artifacts. International Journal of Pattern Recognition and Artificial Intelligence 20 (4), 609–628.
- Zhu, S., Yang, X., 2000. The seamline removing in the generation of orthophoto maps. In: International Archives of Photogrammetry and Remote Sensing. Vol. 33. International Society for Photogrammetry and Remote Sensing, Amsterdam, The Netherlands, pp. 1247–1251.
- Zitová, B., Flusser, J., 2003. Image Registration Methods: A Survey. Image and Vision Computing 21 (11), 977–1000.