

# Puesta en correspondencia de imágenes aéreas y un ortomapa

Sergio Hernández Jiménez

**Resumen** — Para facilitar a los bomberos la extinción de incendios se quiere hacer uso de técnicas de visión por computador con el objetivo de detectar elementos sobre un terreno y geolocalizarlos. Este TFG se basa en una manera de llevar a cabo la geolocalización: Poner en correspondencia las imágenes tomadas por un vehículo aéreo no tripulado (UAV) con un ortomapa del cual ya se conocen las coordenadas de cada pixel. Pero las imágenes y el ortomapa presentan diferencias que dificultan una correlación basada en la detección y emparejamiento de zonas parecidas. Por este motivo no solo queremos crear un sistema que realice la puesta en correspondencia, sino también evaluar las fases de esta puesta en correspondencia y aplicar mejoras con el objetivo de conseguir que las imágenes se vean más parecidas (como por ejemplo reducir el tamaño de las imágenes).

**Palabras clave** — Extinción de incendios, geolocalización, UAV, ortomapa, detección de características, descripción de características, emparejamiento imagen-mapa, igualación de la resolución, espacio escala, pirámide multiresolución.

**Abstract** — To facilitate firefighters their work we want to make use of computer vision techniques with the aim of detect elements on land and geolocate them. This TFG is based on a way of carrying out geolocation: Put in correspondence images taken by an unmanned aerial vehicle (UAV) with an ortomap which coordinates of each pixel are known. But the images and the ortomapa have differences that hinder a correlation based on detect and match similar areas. Therefore we not only want to create a system that performs the put in correspondence, but also evaluate the implementation phases of this correspondence and implement improvements in order to get the images look more similar (such as reduce the size of the images).

**Index Terms** — Firefighting, geolocation, UAV, ortomap, feature detection, feature description, image to map matching, resolution equalization, space scale, multiresolution pyramid.



## 1 INTRODUCCIÓN

Este trabajo de final de grado nace de un objetivo de otro proyecto: el proyecto FireGUARD [1].

Para facilitar la labor de extinción de incendios, el Ministerio de Economía y Competitividad financió el proyecto de investigación FireGUARD, que tiene como finalidad el desarrollo de nuevas tecnologías para ayudar a los bomberos en sus actividades de mitigación de incendios. Se utiliza una flota de vehículos aéreos no tripulados (UAV) equipados con cámaras sensibles al espectro visible e infrarrojo, para tomar imágenes del terreno y poder localizar también frentes de fuego y puntos calientes. Esta información se integrará en un sistema central donde se geolocalizará y donde se coordinarán los recursos de extinción de incendios. También se generará un mapa actualizado del estado del incendio.

Del proyecto general FireGUARD nos interesa el subproyecto FireWATCHER [2], que se centra en utilizar algoritmos de Visión por Computador para, de forma automática, conseguir detectar y caracterizar frentes de fuego y puntos calientes, y poderlos geolocalizar.

Este TFG se sitúa en la parte de geolocalización y se basa

en una posible manera de llevarla a cabo: Poner en correspondencia las imágenes tomadas por un UAV con un ortomapa del cual ya se conocen las coordenadas de cada pixel.

### 1.1 Objetivos

El objetivo global del trabajo es obtener de manera automática la correspondencia entre las imágenes del UAV y un ortomapa. Es decir, determinar para cada punto de una captura cuál es su correspondiente en el ortomapa y viceversa. Con esta correspondencia se podrá:

- Geolocalizar zonas y elementos de las imágenes del UAV, al saber de un punto de esta imagen cual le corresponde en el ortomapa ya geolocalizado.
- Formar un mapa tratando cada imagen independientemente, dando la posibilidad de paralelizar y obtener rápidamente el resultado.

Antes de mostrar los subobjetivos en los que se ha dividido el objetivo global se explican cuestiones que han llevado a ellos y que son necesarias para entenderlos.

#### 1.1.1 Método basado en características

Para realizar la puesta en correspondencia se ha utilizado el método basado en características por ser el más prometedor del estado del arte. Este método, básicamente, detecta puntos de interés (características) en dos pares de

- E-mail de contacto: [sergio.hernandezjm@e-campus.uab.cat](mailto:sergio.hernandezjm@e-campus.uab.cat)
- Mención realizada: Computació.
- Trabajo tutelado por: Daniel Ponsa Mussarra (Ciències de la Computació)
- Curso 2014/15

imágenes, extrae descriptores de estos puntos (vectores que contienen información sobre el entorno del píxel) y empareja los puntos de interés mediante la similitud de sus descriptores. Es necesario decir que existen varios algoritmos para realizar la detección de características y la extracción de descriptores.

En el Estado del Arte se explicará con más detalle el método y otras opciones que se encontraron para realizar la correspondencia que no se han probado.

### 1.1.2 Diferencias entre las capturas y el ortomapa

En un TFG del semestre anterior que trataba el mismo tema [3] se encontraron dificultades para obtener correspondencias correctas a causa, básicamente, de las diferencias entre las imágenes del UAV y el ortomapa. Al estar tomadas las capturas y el ortomapa por cámaras diferentes y en momentos distintos estas plantean diferencias como las siguientes:

- Son de diferente resolución.
- Tienen distintas condiciones de iluminación (sombras, brillos,...).
- Pueden mostrar diferentes elementos en el área observada (objetos nuevos, diferente estado de la vegetación...).

En la Fig. 1 se puede ver una captura tomada por el UAV y su zona equivalente en el ortomapa, y en el apéndice A1 se muestran todas las capturas utilizadas en el trabajo y el ortomapa.

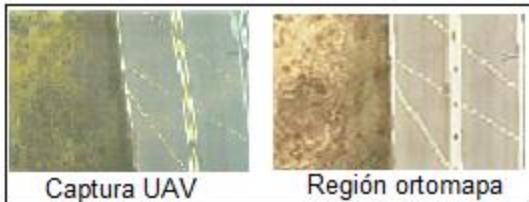


Fig. 1. Captura: 'MyVideo\_14\_58\_58\_219\_062.png' y su región correspondiente del ortomapa.

Estas diferencias dificultan una correlación basada en la detección y emparejamiento de zonas parecidas. Por este motivo se han realizado experimentos para evaluar las diferentes partes de la puesta en correspondencia.

Estos experimentos prueban diferentes detectores y descriptores de características para determinar cuáles funcionan mejor en nuestro caso, pero su objetivo principal es utilizar una técnica que, en principio, hace más parecidas las imágenes y con ella conseguir mejores resultados en la correspondencia.

La técnica consiste en aplicar un espacio escala: Reducir en varios niveles el tamaño de las imágenes del UAV y del ortomapa con la idea de que al disminuir el tamaño, la información importante que se mantiene en los dos pares de imágenes debería ser más parecida.

Antes de aplicar el espacio escala es necesario igualar la resolución de las capturas y del ortomapa para que las imágenes del espacio sean del mismo tamaño.

### 1.1.3 Lista de subobjetivos del proyecto

Tras estas explicaciones previas se muestran a continuación los subobjetivos que se han tratado para lograr el objetivo global de obtener la correspondencia entre las imágenes del UAV y el ortomapa:

1. **Igualar la resolución de las capturas a la del ortomapa automáticamente:** Se aplica un tratamiento inicial al conjunto de imágenes del UAV para solucionar la diferencia de resolución de las capturas con la del ortomapa e intentar conseguir que los píxeles sean más parecidos. Se igualan las capturas a la resolución conocida del ortomapa de manera automática a partir de los datos de calibración de la cámara (o ángulo de visión) y la altura de vuelo.
2. **Crear el espacio escala:** Partiendo de que las capturas y el ortomapa son del mismo tamaño gracias al objetivo anterior, se reescalan a diferentes tamaños inferiores, creando así una pirámide multiresolución.
3. **Aplicar y evaluar detectores y descriptores:** Se aplica y se evalúan adaptando estas fases del método basado en características para poder utilizar la pirámide multiresolución.
4. **Desarrollar un sistema automático image-2-map registration:** Este sistema delimita para cada captura una zona de búsqueda en el ortomapa a partir de información sobre los parámetros de vuelo del UAV en el momento en que tomaba la captura. Entonces realiza la correspondencia entre la captura y la zona de búsqueda mediante la estimación de una homografía [4]. La correspondencia la realiza utilizando el método basado en características en el espacio escala.

En el apéndice A2 se muestra un árbol de objetivos que contiene los objetivos anteriores organizados y agrupados según el orden de su realización.

Para explicar el desarrollo de los objetivos y el resto de cuestiones de interés del artículo, este se ha organizado en los siguientes apartados:

- **Estado del arte:** Expone dos métodos capaces de poner en correspondencia dos imágenes, explica el escogido (el basado en características) y comenta técnicas utilizadas en otros proyectos parecidos al nuestro.
- **Metodología:** En esta sección se muestra la metodología con la que se ha trabajado en el proyecto.
- **Desarrollo del trabajo:** En este apartado se explica el trabajo realizado para cumplir los objetivos. Se indican los experimentos y se analizan los resultados de las evaluaciones de detectores y características y del sistema automático image2map registration.
- **Conclusiones:** Expone las aportaciones realizadas en el proyecto, las conclusiones que extraemos de estas y las posibles líneas de continuación.

## 2 ESTADO DEL ARTE

Se ha buscado información sobre métodos capaces de realizar la correspondencia entre dos pares de imágenes que muestran un mismo lugar, pero que presentan diferencias como las expuestas anteriormente. También se han buscado otros proyectos que realicen una correspondencia en casos parecidos al que nos interesa.

### 2.1 Métodos que ponen en correspondencia imágenes

Se han encontrado dos estrategias principales que realizan una puesta en correspondencia de un par de imágenes: El método directo y el método basado en características [5].

Tanto el método basado en características como el método directo, que correlaciona las imágenes comparando de varias maneras los valores de intensidad de los píxeles, se esperaban probar. Se encontraron códigos a partir de los cuales se podría haber partido tanto para el método directo [6] como para el método basado en características [7], pero por cuestiones de tiempo solo se ha podido probar uno de los métodos.

Se ha escogido el método basado en características al ser el más prometedor encontrándolo en más proyectos parecidos al nuestro, incluido el TFG del semestre anterior [3]. Por ser la técnica que se ha acabado usando se explica a continuación etapa a etapa.

#### 2.1.1 Método Basado en Características

El método basado en características necesita 4 etapas para realizar la correspondencia entre dos imágenes:

- **Detección de características:** Busca en el par de imágenes puntos o zonas que cumplan unas características concretas. Existen diferentes algoritmos (Harris [8], FAST [9], MinEigen [10], MSER [11]...) los cuales determinan el tipo de característica a buscar (básicamente esquinas o regiones) y la manera de buscarla.
- **Extracción de descriptores:** Consiste en aplicar un algoritmo de descripción a las características encontradas en el paso anterior, con el objetivo de poder comparar entre sí las de los dos pares de imágenes y determinar las similitudes. En este caso también existen diferentes algoritmos (BRISK [12], SURF [13]...). Estos generan de diferentes maneras un vector (denominado descriptor local) con información sobre el entorno de una característica.
- **Emparejamiento de características:** Se comparan los descriptores generados en el paso anterior para determinar, de una característica de una imagen, cual de la otra imagen presenta mayor similitud y emparejarlas. Este emparejamiento ya representa una correspondencia entre esas características, pero para obtener una correlación general del par imágenes se realiza el siguiente paso.
- **Estimación de la transformación:** Se estima la transformación a aplicar a un punto de la imagen para obtener el correspondiente en la imagen pareja. Esta estimación de la transformación se denomina homografía y en el momento de estimarla también se utiliza el algoritmo RANSAC [14], que

elimina puntos emparejados muy distantes entre si respecto al resto de emparejamientos.

### 2.2 Proyectos de interés

Los siguientes proyectos crean correspondencias entre imágenes mediante técnicas que no se han probado, de nuevo, por cuestiones de tiempo. No obstante, son interesantes ya que aparecen ideas y métodos que sí probamos, mostrando que esas ideas y métodos se utilizan y son eficaces en situaciones de un ámbito parecido. Básicamente nos referimos al método basado en características y a las ideas de reducción de tamaño y de si es posible poner en correspondencia imágenes que presentan diferencias.

En la presentación de M. Irani y P. Anandan [15] aparece un método directo para correlacionar imágenes de un mismo lugar pero tomadas por diferentes cámaras (sensores), basándose en aplicar un filtro laplaciano para suprimir la información no común y quedarse con la esencial. Nos interesa el hecho de que el par de imágenes que ponen como ejemplo (ver Fig. 2) son fotografías tomadas por diferentes sensores y que se ven bastante diferentes. Esto indica que sí que existen métodos pensados para casos parecidos a nuestro problema.



Fig. 2. Imágenes que se correlacionan en el proyecto de M. Irani y P. Anandan.

En el proyecto de H. Hild [16] se realiza la correspondencia utilizando un método basado en características. Detecta características y las correlaciona pero además, empieza encontrando objetos poligonales en las imágenes a partir de los valores de intensidad de los píxeles, detectando las características en estos objetos.

En el artículo de R. Hardie, K. Barnard i E. Armstrong [17] destaca una manera de mejorar la correlación entre imágenes de diferentes resoluciones. Para la foto de mayor resolución se coge un grupo píxeles y se trata como si fuera uno solo, mirando de reducir el tamaño de este grupo a uno que se adecúe al pixel de la imagen de menor resolución. Esta idea la encontramos equivalente a nuestra reducción de la imagen de mayor resolución.

## 3 METODOLOGÍA

De cara a la metodología seguida para realizar el proyecto, se ha hecho una planificación inicial usando un diagrama de Gantt y un seguimiento semanal, donde se ha ido realizando un análisis de cada objetivo una vez implementado para decidir las siguientes acciones a tomar (ver si es necesario realizar cambios o si se puede seguir según lo planeado).

La planificación ha dividido el trabajo en las siguientes fases. Para las fases que representan objetivos (las fases 3, 4, 5 y 6) se indican los métodos que se han usado y los códigos de los que se ha partido:

#### **Fase 1: Búsqueda de información**

Corresponde al trabajo reflejado en el estudio del arte, donde se han buscado métodos capaces de realizar una correspondencia entre imágenes y otros proyectos interesantes de cara al nuestro.

#### **Fase 2: Preparación del entorno de trabajo para los objetivos**

En esta fase se ha concretado la tecnología a utilizar y las herramientas de trabajo, y se ha realizado la instalación del programario Matlab. Se ha escogido Matlab por disponer de un amplio abanico de librerías de tratamiento de imágenes, que incluyen funciones para realizar las etapas del método basado en características.

#### **Fase 3: Igualación automática de la resolución**

Se igualan automáticamente las resoluciones de las capturas a la del ortomapa haciendo uso de las relaciones entre los elementos de una adquisición de imágenes. Con este método se obtiene el factor escala a aplicar a una captura para reducirla al tamaño del ortomapa.

#### **Fase 4: Creación del espacio escala**

Reducimos el tamaño de las imágenes aplicando un factor escala que varía proporcionalmente según cada nivel del espacio creando así una pirámide multiresolución.

#### **Fase 5: Aplicación y evaluación de detectores y descriptores**

Se lleva a cabo partiendo de un código [7] que realiza una correspondencia entre una imagen y la misma imagen alterada (rotada y disminuida de tamaño), y del código del TFG del semestre pasado que trataba el mismo tema [3]. Se han tenido que adaptar para poder utilizar el espacio escala.

#### **Fase 6: Desarrollo del sistema automático image-2-map**

Para cada captura su zona aproximada del ortomapa se recorta según un área (que mide tantos metros como se especifique) centrada a partir de datos sobre la geolocalización del UAV. Para la puesta en correspondencia se parte de la aplicación de detectores y descriptores de la fase anterior (se escogen los mejores detectores y descriptores según los experimentos). Se completa las partes que faltaban del método basado en características (el emparejamiento y la estimación de la transformación) utilizando de nuevo el código de [7].

## **4 DESARROLLO DEL TRABAJO**

En este apartado se muestra el trabajo que se ha realizado para cumplir los objetivos. Primero se habla del dataset con el que se ha trabajado y se explica la igualación automática de las resoluciones de las capturas a la del ortomapa. Posteriormente se indican los experimentos realizados para las evaluaciones de detectores, descriptores y el sistema automático image2map y se analizan sus resultados.

### **4.1 Dataset**

Para poder realizar los experimentos se ha partido del dataset del TFG del semestre anterior [3]. En el dataset hay un conjunto de 10 imágenes extraídas de un video realizado por un UAV. Este video se tomó durante un vuelo sobre el campo de aviación de Sant Cugat del Vallés. El dataset también está formado por un ortomapa del campo de aviación obtenido del *Institut Cartogràfic de Catalunya*, cuya resolución es conocida: 25 cm/pixel. Las resoluciones de las capturas no las sabemos directamente y es necesario calcularlas para conocerlas, pero por sus tamaños en píxeles sabemos de entrada que son mayores a la del ortomapa.

Las capturas al sacarlas del video están en color y tienen un tamaño de 720x576 píxeles. No obstante, a causa de ser tomadas en movimiento y por el doble barrido de captura de la cámara, sufrían de un efecto conocido como efecto peine que causaba una distorsión en la imagen. Se solucionó en el TFG anterior [3] extrayendo las columnas y filas impares que provocaban redundancia y que causaban la distorsión.

También se extrajeron bordes oscuros que presentaban las imágenes y, con todo esto, el tamaño de las capturas con las que trabajamos quedó reducido a 336x287 píxeles.

En el apéndice A1 aparecen las capturas del UAV y el ortomapa utilizado.

### **4.2 Igualación automática de la resolución**

La idea es igualar la resolución de las capturas a la del ortomapa (ya que las capturas son de mayor resolución) buscando conseguir que los pixeles de ambas se parezcan más.

Esta igualación se realiza de manera automática, de forma diferente al TFG del semestre anterior [3], que también igualaba la resolución pero utilizando una anotación manual de puntos de correspondencia para cada captura y el ortomapa. El objetivo de nuestra igualación es que solo necesite datos de la cámara y no una anotación manual.

A partir de las resoluciones de una captura y del ortomapa se puede obtener el factor escala necesario para reescalcar la captura al tamaño de su pareja (dividiendo las dos resoluciones).

La resolución del ortomapa es de 25 cm/pixel, pero las de las capturas las tenemos que calcular. Como sabemos las dimensiones en píxeles de las imágenes y la resolución la tratamos en cm/pixel, para obtener la resolución necesitamos el tamaño real de las áreas que las capturas contienen.

Por suerte, disponemos de la focal de la cámara que sacó el video y de la altura en la que se encontraba el UAV en el momento de capturar cada frame.

Entonces, a partir del tamaño en píxeles de la imagen tomada por el sensor de la cámara (S), de la focal (V) y de la altura (U), podemos utilizar las relaciones entre estos elementos (1) para obtener el tamaño del área (FOV). Se pueden ver estos elementos en la Fig. 3.

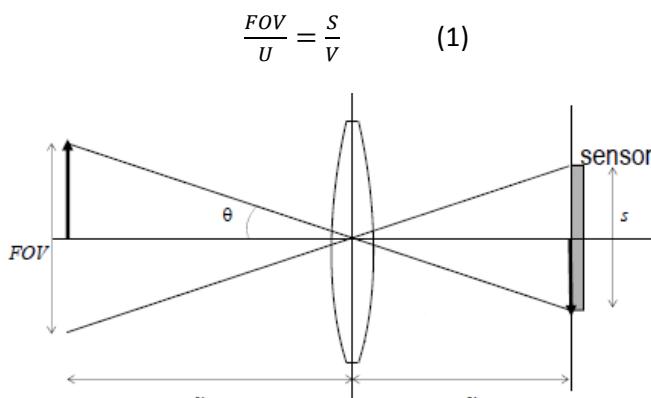


Fig. 3. Elementos en una adquisición de imágenes.

De esta manera obtenemos primero el tamaño del área, después la resolución de la captura y, finalmente, el factor escala a aplicar a la captura para conseguir reescalarla a la resolución del ortomapa.

#### 4.3 Creación de la pirámide multiresolución

Queremos disminuir en varios niveles el tamaño de las capturas del UAV y del ortomapa creando una pirámide multiresolución [18]. La motivación de esto es que la técnica consiga hacer más parecidas las imágenes y mejore los resultados del método basado en características.

También es importante indicar que, para la realización de experimentos, se prepara la creación del espacio escala de manera que sea capaz de obtener diferentes pirámides según unos parámetros.

Gracias al punto anterior ahora las capturas del UAV y el ortomapa tienen la misma resolución. Entonces, para reducir el tamaño de las imágenes, se aplica un factor escala que varía proporcionalmente según cada nivel de la pirámide (en el primer nivel se conserva el tamaño original).

Por ejemplo, si el ancho de una imagen es de 500 píxeles y el factor escala es de 0.8, el ancho de la imagen de primer nivel será de 500 píxeles, el de segundo nivel será de 400 (500\*0.8), el de tercer nivel de 320 (400\*0.8), etc. Se van haciendo niveles hasta que el ancho o el alto de la imagen alcanzan o sobrepasan un tamaño mínimo fijado.

Esta reducción de tamaño según un factor la podemos realizar con dos alternativas:

- Utilizar el factor 0.5 y suavizar con una gaussiana de sigma 2.
- Utilizar un factor próximo a 1 y reescalar con interpolación bicúbica sin suavizar.

#### 4.4 Motivación de la evaluación de detectores y descriptores

Todas las partes del método basado en características se basan en la idea de que las zonas equivalentes en dos imágenes deberían ser más parecidas entre ellas que entre cualquier otra zona. De esta manera se detectarían características en los mismos lugares de las dos imágenes, los descriptores serían más parecidos y el emparejamiento daría buenos resultados.

Pero las diferencias que presentan las capturas del UAV respecto al ortomapa nos hacen dudar de que esta idea se

cumpla lo suficientemente bien en nuestro caso. Por este motivo ha sido necesario realizar experimentos para determinar de qué manera funciona mejor el método en nuestro dataset.

Se prueban algunos detectores y descriptores pero la comprobación más importante que se quiere hacer es si el espacio escala consigue que las imágenes sean más parecidas en tamaños más pequeños y, con esto, mejorar los resultados.

#### 4.5 Aplicación y evaluación de detectores

Se hace una evaluación del proceso de detección de características en base a determinar el porcentaje de puntos detectados en una captura y en su zona correspondiente del ortomapa que coinciden.

Seguidamente se muestran los pasos de la evaluación en términos generales para una captura. Más adelante se indican los detectores usados, así como los parámetros de estos y de la generación del espacio.

1. Se convierten la captura y el ortomapa a escala de grises.
2. Se iguala la resolución de la captura a la del ortomapa con el método automático creado.
3. Se obtiene la región del ortomapa equivalente a la captura aplicando una homografía generada con una anotación manual de puntos de correspondencia entre la captura y el ortomapa. Las homografías las hemos obtenido del TFG que trabajaba el mismo tema [3].
4. Se genera una pirámide multiresolución para la captura y otra para la región equivalente del ortomapa.
5. Para cada nivel de la pirámide se aplica el detector sobre la captura y la región del ortomapa.
6. Se selecciona la imagen de estas dos donde se han detectado menos puntos.
7. Para obtener el resultado de la evaluación se calcula el porcentaje de puntos de la imagen seleccionada que tiene algún punto de la otra a una distancia de 0, 1, 2, 3 o 4 píxeles. No nos fijamos solo en el porcentaje de puntos con distancia 0 ya que, al utilizar una homografía para obtener la zona del mapa correspondiente a una captura, es probable que todos los píxeles de las dos imágenes no coincidan perfectamente.

##### 4.5.1 Parámetros para el espacio escala

Respecto a los parámetros del espacio escala, en primer lugar se probó la opción que creaba la pirámide con el factor de escala 0.5, pero se vio que el tamaño inicial de las imágenes no era demasiado grande y al cabo de muy pocos niveles se llegaba a un tamaño de imagen bastante pequeño, sin que los detectores encontraran puntos en varias ocasiones. Por lo tanto se siguió con la otra opción de pirámide, a la que le indicamos un factor cercano a 1.

Haciendo varias pruebas con la evaluación se ha observado que se obtienen buenos resultados en los porcentajes de puntos con un factor escala de 0.7 y un tamaño mínimo de 32 píxeles, y son los que nos hemos quedado para la generación de los resultados que analizamos.

#### 4.5.2 Detectores y parámetros utilizados

De cara a los detectores utilizados, se han hecho evaluaciones con los detectores MinEigen, Harris y FAST, al ser los que iban mejor en el TFG que trataba el mismo tema [3].

En la Tabla 1 se muestra el valor de los parámetros (columnas) que se han acabado usando con cada detector (filas).

	<b>FilterSize</b>	<b>MinQuality</b>	<b>MinContrast</b>
<b>MinEigen</b>	3	0.1	No se usa
<b>Harris</b>	3	0.01	No se usa
<b>FAST</b>	No se usa	0.2	0.05

Tabla 1. Valores de los parámetros usados en la detección de características.

#### 4.5.3 Ejemplo de detección de características

Para mostrar un ejemplo de la detección, en la Fig. 4 aparece el nivel de menor tamaño de una de las capturas del dataset y su zona equivalente del ortomapa con los puntos detectados. Se ha realizado esta detección usando el detector Harris y el resto de parámetros indicados en los subapartados anteriores.

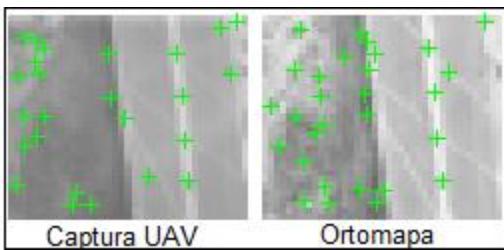


Fig. 4. Detección de puntos con detector Harris (captura: MyVideo14\_58\_58\_219\_062.png, nivel de la pirámide: 4).

En las imágenes de la Fig. 4 observamos que en la parte de tierra de la izquierda el terreno es diferente en ambas imágenes, y por tanto no se aprecia un gran número de coincidencias. Lo bueno es que en la zona de la pista (que no ha cambiado de una imagen a otra) sí que se ven bastantes puntos detectados en la captura y en el ortomapa que coinciden.

#### 4.5.4 Tabla de resultados

Se han agrupado los resultados de la evaluación en una tabla para poder estudiarlos mejor. A continuación se indica la estructura de esta tabla, se muestra la tabla y se analizan los resultados. La Tabla 2 contiene los siguientes datos:

- Detector: El nombre del detector evaluado.
- Nivel: El número del nivel de la pirámide multi-resolución. Este se incrementa a medida que se va reduciendo el tamaño de las imágenes y el nivel 1 representa el tamaño original.
- %D0 - %D4: Representan los porcentajes de puntos detectados coincidentes en la captura y el ortomapa, a una distancia de 0, 1, 2, 3 y 4 píxeles (Para un nivel y un detector). Cada casilla contiene la media de todas las capturas del dataset para ese caso concreto.

<b>Detector</b>	<b>Nivel</b>	<b>%D0</b>	<b>%D1</b>	<b>%D2</b>	<b>%D3</b>	<b>%D4</b>
MinEigen	1	23.29	49.19	62.14	69.52	78.25
MinEigen	2	27.38	54.27	67.76	76.80	83.51
MinEigen	3	32.09	55.16	70.45	82.94	91.49
MinEigen	4	38.29	56	72.10	81.85	84.03
Harris	1	22.12	49.44	64.32	75.49	80.45
Harris	2	26.42	54.65	75.58	85.89	92.37
Harris	3	33.97	61.20	77.42	88.43	94.76
Harris	4	40.28	67.82	79.71	88.80	94.84
FAST	1	7.10	42.67	69.35	78.96	84.16
FAST	2	13.56	54.63	76.56	87.05	90.73
FAST	3	16.62	59.63	79.53	86.02	90.03
FAST	4	32.38	69.21	85.00	90.77	92.57

Tabla 2. Resultado de los experimentos del detector Harris (captura: MyVideo14\_58\_58\_219\_062.png).

Lo más importante que vemos en la tabla es que existe una tendencia en el aumento de porcentajes de puntos coincidentes a medida que se avanza de nivel en la pirámide. Esto indica que en imágenes de inferior tamaño se encuentran más puntos en los mismos lugares de las capturas y del ortomapa, y que por tanto, el espacio escala ha funcionado de manera efectiva en la detección de características.

Respecto a las distancias de píxeles consideradas, quizás para puntos de una misma imagen bastantes cercanos, una distancia de 3 píxeles o 4 ya indica otro punto distinto al que se mide, pero una distancia de 2 píxeles se puede considerar bastante correcta, y porcentajes alrededor del 85% en los niveles 4 parecen un buen resultado.

Finalmente, de cara al detector a escoger como mejor, se optaba entre Harris y FAST, por obtener generalmente mejores resultados que MinEigen en niveles avanzados de la pirámide. Se ha acabado eligiendo Harris ya que consigue más coincidencias en la distancia de 0 píxeles.

#### 4.6 Aplicación y evaluación de descriptores

Los puntos detectados en el paso anterior se pasan a un descriptor y son utilizados para la evaluación.

Los primeros pasos para poder realizar la evaluación de los descriptores son los mismos que los pasos del 1 al 6 de la evaluación de detectores. A continuación se muestran los pasos propios de esta evaluación, que siguen a los otros:

7. Se aplica un descriptor a la captura y a la zona equivalente del ortomapa para describir los puntos de la captura o de la zona del ortomapa (se escogen los puntos de la imagen donde se han detectado menos y estos se utilizan en las dos imágenes para que la descripción sea sobre lo mismo).
8. Para obtener el resultado de la evaluación se generan las distancias de todas las descripciones de puntos de una imagen con todas las de la otra, y se calcula el porcentaje de puntos cuya distancia mínima corresponde a la de su punto homónimo en la otra imagen.

#### 4.6.1 Parámetros y descriptores utilizados

De cara a los valores de parámetros para el espacio escala se han usado los mismos que en la evaluación de detectores.

Respecto a los algoritmos de descripción utilizados, primero se probaron BRISK, FREAK y SURF. Pero a causa de un filtrado de puntos que realizan los algoritmos, con BRISK y FREAK apenas quedaban puntos descriptos y en muchos casos no quedaba ningún punto. Esto evitaba obtener porcentajes de puntos fiables y por este motivo en la tabla de resultados solo mostramos los de SURF.

#### 4.6.2 Tabla de resultados

Los resultados de la evaluación se muestran en la Tabla 3 donde se ha probado el descriptor SURF con los puntos resultantes de los 3 detectores. Los datos de la tabla son los siguientes:

- Detector: El nombre del detector evaluado.
- Nivel: El número del nivel de la pirámide multi-resolución.
- %Desc. SURF: Representan el porcentaje de puntos cuya distancia mínima en el espacio de descripción corresponde a la de su punto homólogo en la otra imagen.

Detector	Nivel	% Desc. SURF
MinEigen	1	25
MinEigen	2	40
MinEigen	3	38
MinEigen	4	52
Harris	1	29
Harris	2	36
Harris	3	48
Harris	4	49
FAST	1	29
FAST	2	40
FAST	3	48
FAST	4	54

Tabla 3. Resultado de los experimentos del descriptor SURF con puntos detectados con Harris (captura: MyVideo\_14\_58\_58\_219\_062.png; SURFSize: 128).

Observamos que se mantiene la misma tendencia que en la evaluación de detectores: el aumento de porcentajes a medida que se avanza de nivel en la pirámide. Esto muestra que el espacio escala también funciona bien en la descripción. Consigue que los descriptores de un punto en una captura y en el ortomapa sean más parecidos.

No obstante, los porcentajes no son altos. Tan solo se llega hasta un 54% de coincidencias (en el nivel 4 y con puntos del detector FAST). Un motivo de esto quizás sea la existencia de puntos parecidos en una misma imagen (por ejemplo las esquinas de dos rayas blancas de la pista). Puede que haya algún punto de estos similares que tenga

un descriptor más parecido en la imagen pareja que el correspondiente al punto que se está analizando.

#### 4.6.3 Eliminación de puntos detectados

Para mejorar los resultados se ha probado una técnica para evitar introducir en el análisis estos puntos que no son los mismos pero que se parecen en cuanto a sus descriptores. La técnica consiste en eliminar según un valor umbral descriptores muy parecidos en una misma imagen.

Se han hecho experimentos con varios valores umbráles, obtenidos de la diferencia entre los descriptores de distintos puntos que se ven parecidos en una imagen. Pero los porcentajes han sido peores que los anteriores y, ya que la eliminación hacia que quedasen aún menos puntos, esta técnica no se ha usado para los resultados de la tabla.

### 4.7 Sistema automático image2map registration

Se ha creado un sistema para poner en correspondencia imágenes del UAV con el ortomapa. Delimita para cada captura una zona de búsqueda aproximada en el ortomapa a partir de datos de la cámara del UAV y aplica el método basado en características con las mejoras creadas en los objetivos anteriores. Se utilizan los algoritmos de detección y descripción escogidos en la evaluación como mejores para nuestro caso (el detector Harris y el descriptor SURF).

#### 4.7.1 Recorte del ortomapa

En la evaluación de detectores y descriptores se obtenía la zona del ortomapa equivalente a una captura a partir de una homografía generada con una anotación manual de puntos. Al igual que pasaba en la igualación de la resolución de las capturas a la del ortomapa, se ha querido hacer el recorte automáticamente.

Utilizando datos de la localización del UAV se obtiene la posición, en función del ortomapa, de los 4 extremos de cada captura (de manera aproximada). A partir de los extremos se calcula el punto del ortomapa correspondiente al punto medio de las capturas. Entonces recortamos del ortomapa un área de 40x40 metros alrededor de ese punto. De esta manera obtenemos una zona más grande que la mostrada por la captura y que la contiene, dado que la localización del UAV puede contener errores. En la Fig. 5 se muestra una de las capturas y el recorte del mapa para esta captura.

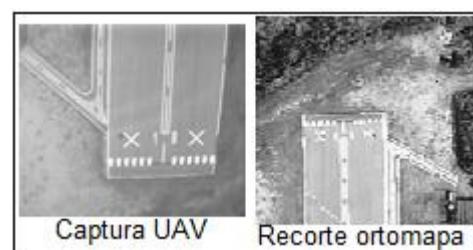


Fig. 5. Captura: MyVideo\_14\_58\_06\_234\_061 y su recorte del ortomapa.

#### 4.7.2 Aplicación del método basado en características

Para obtener las correspondencias entre las capturas del UAV y el ortomapa se utilizan las siguientes fases:

1. Se convierten las capturas y el ortomapa a escala de grises.

2. Para cada captura se iguala su resolución a la del ortomapa con el método automático creado.
3. Se recorta la región del ortomapa correspondiente a la captura.
4. Se genera una pirámide multiresolución para la captura y otra para el recorte del ortomapa.
5. Para cada nivel de la pirámide se aplica el detector Harris sobre la captura y el recorte del ortomapa, acumulando para una captura los puntos detectados en todos sus niveles de la pirámide.
6. Se proyectan los puntos de todos los niveles sobre la captura y el recorte del ortomapa de nivel 1, y se aplica SURF para extraer los descriptores.
7. Se emparejan los puntos proyectados sobre la captura y el recorte del ortomapa según el emparejamiento del método basado en características.
8. Se calcula la homografía entre la captura y el ortomapa a partir del emparejamiento anterior.

#### 4.7.3 Análisis del emparejamiento

Uno de los objetivos iniciales de este TFG era hacer una evaluación del sistema automático image2map registration que aportase números y porcentajes (como las evaluaciones de detectores y descriptores de características), pero se acabó el tiempo y no se pudo realizar. No obstante, se han hecho algunas pruebas simples para hacernos una idea de cómo de bien el sistema realiza la correspondencia entre los puntos detectados de las imágenes (como de bien los empareja). Para la prueba se muestran las capturas y los recortes del ortomapa con el emparejamiento de puntos encima.

Para las funciones que realizan el emparejamiento y la homografía se han tenido que probar valores de varios parámetros, ya que con los valores por defecto no se generaban emparejamientos. Estos parámetros, generalmente, se refieren a lo parecido que tienen que ser los puntos para considerarlos el mismo y emparejarlos. Por tanto, se han ajustado de manera que sean más permisivos y generen más emparejamientos. Los valores que finalmente se han escogido para los parámetros de las funciones son los siguientes:

- **matchFeatures**('MatchThreshold':20, 'MaxRatio':0.8).
- **estimateGeometricTransform**('transformType':projective, 'MaxDistance':25).

El hecho de que los valores por defecto de los parámetros no generasen emparejamientos ya indicaba que los puntos (sus descriptores concretamente) no eran muy parecidos. Y se ha visto en las pruebas que, después de aceptar con los parámetros parejas de puntos que sean más diferentes, los emparejamientos de estos puntos no son correctos en la mayoría de los casos.

En la Fig. 6 se muestra un ejemplo del emparejamiento para una captura y un recorte del ortomapa.

Se ve un mal emparejamiento, por ejemplo, en el punto detectado en el bosque en el ortomapa que se ha detectado en la captura en una línea de la pista en la captura (línea en verde), y también en el punto detectado en una línea lateral de la pista en el ortomapa que se ha detectado en la línea

central de la pista en la captura (línea en amarillo).

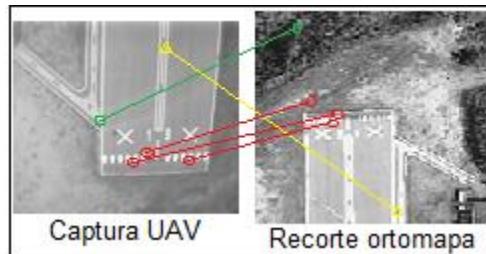


Fig. 6. Emparejamiento de puntos entre la captura: MyVideo\_14\_58\_06\_234\_061 y su recorte del ortomapa.

Después de los buenos resultados en la evaluación de detectores, que no se conseguían igualar en el caso de descriptores, parecía que el principal motivo de esto eran los puntos similares en una misma imagen (que provocarían confusiones a la hora de emparejar), pero no nos aseguraba que los descriptores de los puntos detectados en los pares de imágenes no fuesen parecidos, que las imágenes fueran bastante diferentes a nivel del entorno del pixel.

Lo que se ha visto del sistema image-2-map confirmaría la idea de los puntos similares de una misma imagen y también indica que no hay puntos detectados en las capturas y en el ortomapa cuyos descriptores se parezcan demasiado.

## 5 CONCLUSIONES

Durante este trabajo de final de grado se han realizado las siguientes aportaciones y se han llegado a las siguientes conclusiones:

- Se ha creado un módulo que iguala la resolución de las capturas del UAV a la del ortomapa de manera automática, sin tener que clicar manualmente puntos de correspondencia entre las imágenes, usando datos del UAV en su lugar.
- Se ha aplicado un espacio escala en las evaluaciones de detectores y descriptores de características y en el sistema automático image-2-map, y se ha visto que ha conseguido mejorar los resultados de todas las evaluaciones. Se concluye que, la idea de que al avanzar de nivel en el espacio escala los porcentajes aumentarían a causa de que la información importante que se mantiene debería ser más parecida entre los dos pares de imágenes, ha sido correcta. Al haber funcionado bien aquí, se encuentra interesante de cara a otros proyectos en los cuales sea necesario comparar imágenes de una misma escena pero que presentan diferencias.
- Se han evaluado los detectores MinEigen, Harris y FAST y en los experimentos se ha concluido que Harris es el mejor para nuestro caso. Se ha observado que la pirámide multiresolución ha conseguido mejoras en los resultados de los detectores del 15% entre los resultados del tamaño original y los del nivel más bajo en muchos casos (en unos pocos casos se ha llegado a mejoras más altas, del 80% incluso).
- Se han evaluado los algoritmos de extracción de descriptores BRISK, FREAK y SURF, y se ha visto que

SURF es el único que aceptaba suficientes puntos detectados y que generaba suficientes descriptores. Con la pirámide multiresolución se han conseguido mejoras de los resultados de alrededor del 45% también entre los resultados del tamaño original y los del nivel más bajo.

- Se ha ideado un mecanismo para recortar del ortomapa una zona aproximada a la mostrada en una captura, también de manera automática con datos del UAV, sin necesidad de anotar manualmente puntos de correspondencia entre las imágenes.
- Se ha desarrollado un software que implementa los pasos necesarios para emparejar imágenes del UAV con el ortomapa basándonos en puntos característicos. Pero tras analizarlo se ha visto que los descriptores de los puntos a emparejar no son parecidos y que los emparejamientos creados no son correctos.
- A partir de los experimentos en las evaluaciones de detectores se ha llegado a la conclusión de que el espacio escala consigue que las imágenes se vean a nivel de pixel lo bastante parecidas como para que se detecten un buen número de puntos en los mismos sitios. No obstante, la evaluación de descriptores y el análisis del sistema image2map apuntan a que no se consigue el suficiente parecido como para lograr descriptores similares que generen un buen emparejamiento.

## 5.1 Líneas de continuación

La principal línea para continuar con el trabajo del proyecto debería basarse en conseguir buenos emparejamientos en el dataset actual.

Se podría empezar realizando una evaluación del sistema automático image-2-map más exhaustiva, que diese resultados numéricos, para obtener más información sobre la diferencia entre los descriptores de la captura y del ortomapa relativos a un mismo punto, el problema que parece evitar que se generen buenos emparejamientos.

Entonces se podría mirar de mejorar la eliminación de descriptores para incluirla en el sistema image-2-map mirando de obtener un mejor valor umbral o realizar la eliminación con algún otro criterio.

Las imágenes se han evaluado en escala de grises. También estaría bien probar las evaluaciones de detectores y descriptores y el sistema image-2-map con las imágenes en su color original, para ver si la mayor dimensionalidad del espacio RGB ayuda a que los descriptores de los puntos parejos sean más parecidos.

Después de esto se podrían probar el método directo o alguna de las otras técnicas que aparecen en el estado del arte, para ver si alguna funciona bien en nuestro caso.

Finalmente, cuando se consiguieran emparejamientos aceptables, se debería hacer el paso de geolocalizar zonas y elementos de las imágenes del UAV y crear un mapa de la zona a partir de las capturas, para acercarnos más a los objetivos del proyecto FireGUARD.

## AGRADECIMIENTOS

Le doy las gracias a Daniel Ponsa, mi tutor en este trabajo de final de grado, por sus consejos semana tras semana y sus exhaustivas correcciones y comentarios de los informes y del trabajo en general.

También agradezco a Felipe Lumbreras sus recomendaciones durante las reuniones.

## BIBLIOGRAFÍA

- [1] FireGuard Project (Lab-UAS). Technical Innovation Cluster in Aeronautical Management, 2013 [consulta 15/06/2015]. Disp. en: [http://centresderecerca.uab.cat/aeronauticalcluster/content/fireguard-project-lab-uasReferència\\_2](http://centresderecerca.uab.cat/aeronauticalcluster/content/fireguard-project-lab-uasReferència_2)
- [2] FireWATCHER, Fire Warning by Aerial Terrain Control of Hot Embers Regions. Centre de Visió per Computador - Advanced Driver Assistance Systems Group, 2015 [consulta 15/06/2015]. Disp. en: <http://www.cvc.uab.es/adas/projects/firewatcher/>
- [3] MARTÍNEZ, Alberto. Puesta en correspondencia de imágenes aéreas y un ortomapa, EE/UAB TFG Informática, 2015.
- [4] DUBROFSKY, Elan. Homography estimation. 2009. Tesis Doctoral. UNIVERSITY OF BRITISH COLUMBIA (Vancouver).
- [5] SZELISKI, Richard. Image alignment and stitching: A tutorial. Foundations and Trends® in Computer Graphics and Vision, 2006, vol. 2, no 1, p. 1-104.
- [6] Intensity based automatic image registration, The MathWorks Inc., R2014b Documentation, 2014. [consulta 16/06/2015]. Disp. en: <http://es.mathworks.com/help/images/intensity-based-automatic-image-registration.html>
- [7] Find image rotation and scale using automated feature matching, The MathWorks Inc., R2014b Documentation, 2014. [consulta 16/06/2015]. Disp. en: <http://es.mathworks.com/help/vision/examples/find-image-rotation-and-scale-using-automated-feature-matching.html>
- [8] HARRIS, Chris; STEPHENS, Mike. A combined corner and edge detector. En Alvey vision conference. 1988. p. 50.
- [9] FAST Algorithm for Corner Detection. OpenCV 3.0.0-dev documentation, OpenCV, 2011-14 [consulta 16/06/2015]. Disp. en: [http://docs.opencv.org/3.0-beta/doc/py\\_tutorials/py\\_feature2d/py\\_fast/py\\_fast.html](http://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_fast/py_fast.html)
- [10] SHI, Jianbo; TOMASI, Carlo. Good features to track. En Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on. IEEE, 1994. p. 593-600.
- [11] MATAS, Jiri, et al. Robust wide-baseline stereo from maximally stable extremal regions. Image and vision computing, 2004, vol. 22, no 10, p. 761-767.
- [12] LEUTENEGGER, Stefan; CHLI, Margarita; SIEGWART, Roland Y. BRISK: Binary robust invariant scalable keypoints. En Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011. p. 2548-2555.
- [13] BAY, Herbert, et al. Speeded-up robust features (SURF). Computer vision and image understanding, 2008, vol. 110, no 3, p. 346-359.
- [14] DERPANIS, Konstantinos G. Overview of the RANSAC Algorithm. York University, Toronto, Canada, 2010.
- [15] IRANI, Michal; ANANDAN, P. Robust multi-sensor image alignment. EnComputer Vision, 1998. Sixth International Conference on. IEEE, 1998. p. 959-966.
- [16] HILD, Heiner. Automatic image to map registration of remote

- sensing data. 2001.
- [17] HARDIE, Russell C.; BARNARD, Kenneth J.; ARMSTRONG, Ernest E. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *Image Processing, IEEE Transactions on*, 1997, vol. 6, no 12, p. 1621-1633.
- [18] SUCAR, L. Enrique; GÓMEZ, Giovani. Visión Computacional, p. 96-98. Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México, 2011.

## APÉNDICE

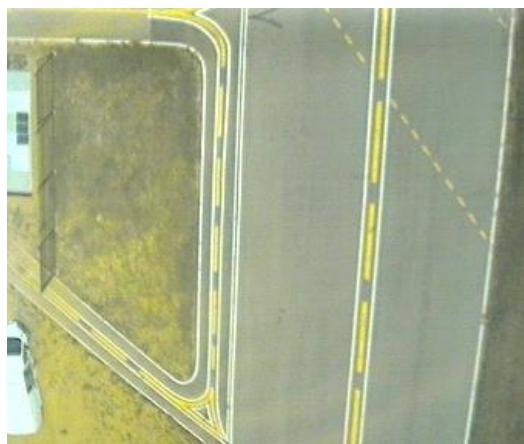
### A1. Imágenes del dataset



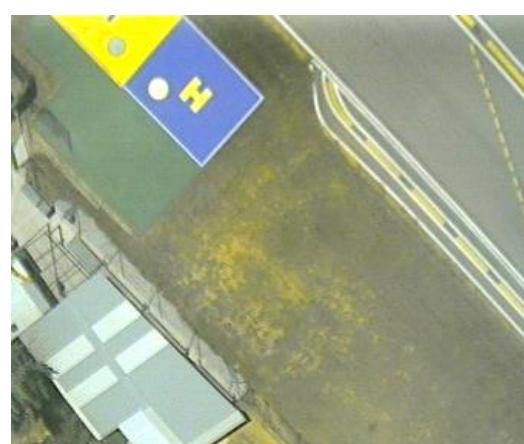
MyVideo\_14\_58\_06\_234\_061.png



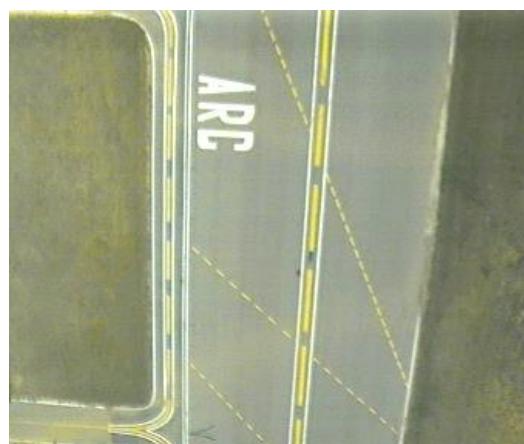
MyVideo\_14\_58\_21\_172\_062.png



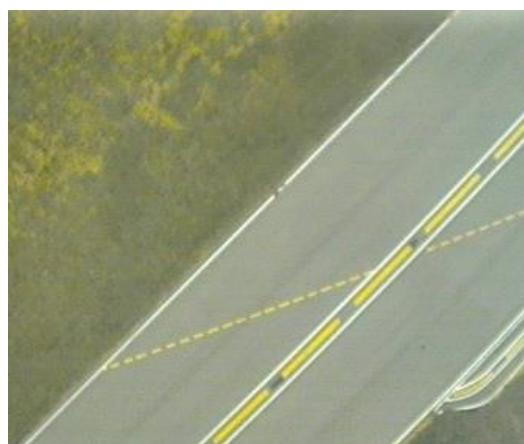
MyVideo\_14\_58\_08\_670\_063.png



MyVideo\_14\_58\_29\_389\_062.png



MyVideo\_14\_58\_17\_391\_063.png



MyVideo\_14\_58\_37\_096\_064.png



MyVideo\_14\_58\_45\_420\_-938.png



MyVideo\_14\_59\_06\_686\_061.png



MyVideo\_14\_58\_58\_219\_062.png



MyVideo\_14\_59\_02\_109\_061.png



CampVolSantCugatWGS84.tif (ortomapa)

## A2. Árbol de objetivos

