

Puesta en correspondencia de imágenes aéreas y un ortomapa

Alberto Martínez Rodríguez

Resumen—Para luchar contra los desastres naturales, especialmente los incendios forestales, se quiere hacer uso de las técnicas de visión por computador y las nuevas tecnologías en general de modo que se pueda geolocalizar elementos detectados sobre el terreno. Una posible manera de realizar esto es determinar a qué zona o punto geográfico corresponde una imagen tomada por un vehículo aéreo no tripulado (UAV), emparejando esta imagen con un ortomapa del cual ya sabemos las coordenadas de cada pixel. Para ello se quiere evaluar si la técnica estándar usada para emparejar dos imágenes sirve para emparejar una imagen con un ortomapa. En este artículo se describe el estado del arte actual de los detectores y descriptores de características locales, describiendo y posteriormente evaluando los más utilizados. Por último se analizan los resultados obtenidos y se determina, tanto para los detectores como para los descriptores, cual se adapta mejor al problema planteado.

Palabras clave—Georeferenciación de imágenes, emparejamiento imagen-mapa, BRISK, FAST, Harris, MinEigen, MSER, SURF, FREAK, SURF, Block, UAV, ortomapa, emparejamiento de imágenes, detección de puntos de interés, descripción de puntos de interés.

Abstract— To combat natural disasters, especially forest fires, we want to use computer vision techniques and the new technologies in general in order to geolocate ground elements. One way to do it is determine which area or geographic point corresponds an image taken by an unmanned aerial vehicle (UAV), matching this image with an ortomapa which we know the coordinates of each pixel. For this purpose we want to review if the standard technique used to match two images allow to match an image with an ortomapa. This paper describes the local features detectors and descriptors state of the art, describing and subsequently evaluating the most used. Finally the results are analyzed and determined for both detectors as descriptors, which is better suited to the proposed problem.

Index Terms—Image georeferentiation, image to map matching, BRISK, FAST, Harris, MinEigen, MSER, SURF, FREAK, SURF, Block, UAV, ortomapa, image matching, interest point detection, interest point description.

1 INTRODUCCIÓN

Actualmente los incendios forestales son un gran problema para muchos países. Debido al impacto del cambio climático, se espera que el peligro de incendios aumente en toda Europa, especialmente en la región mediterránea ya propensa a padecerlos. Los impactos negativos de los incendios forestales son muchos: dañan el ecosistema, causan enormes pérdidas económicas, y ponen a las personas y propiedades en riesgo, entre otros. Por otra parte, la larga vida de los árboles hace de la recuperación de los bosques un proceso muy lento, por lo que procurar su preservación es una tarea de gran importancia.

En este contexto se creó FireGUARD [1], un proyecto de investigación financiado por el Ministerio de Economía y Competitividad que tiene como objetivo el desarrollo de nuevas tecnologías orientadas a preservar las masas forestales. En concreto, el objetivo es desarrollar un sistema para ayudar a los bomberos en misiones de extinción, para aumentar la eficacia de sus actividades de mitigación de incendios. Por medio de una flota de vehículos

aéreos no tripulados (UAV) (ver apéndice A1), se generará de forma continua un mapa actualizado del estado del incendio. Los UAVs estarán equipados con cámaras sensibles al espectro visible e infrarrojo, que transmitirán sus resultados a una estación terrestre. En esta estación las imágenes se procesarán a fin de que se localice el frente de fuego y los puntos calientes (brasas), cuantificando los atributos relevantes de interés (zona, tamaño de la llama, etc.). Esta información se integrará en un sistema de información geográfica, donde se centralizará también la situación de los recursos de extinción de incendios que se encuentren trabajando sobre el terreno. En base a estos datos, un sistema de gestión de flotas coordinará los recursos de tierra para optimizar la misión de extinción de incendios y evitar situaciones de alto riesgo.

El proyecto FireGUARD alberga tres subproyectos: FireMANIA, FireWATCHER y FireFIGHT. El que nos interesa es el segundo.

El subproyecto FireWATCHER [2] se centra en el uso de algoritmos de Visión por Computador para procesar los datos proporcionados por la flota de UAV. Se aplicarán técnicas de este campo para detectar y caracterizar los puntos calientes y frentes de fuego en las imágenes aéreas adquiridas de forma remota. Se necesitarán algoritmos

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

específicos para extraer la información necesaria de forma automática y georreferenciar con precisión los elementos detectados.

En concreto en el TFG “Puesta en correspondencia de imágenes aéreas y un ortomapa” se implementarán y evaluarán diferentes técnicas para poner en correspondencia secuencias de imágenes aéreas con un ortomapa con el objetivo de georreferenciar las imágenes tomadas por un UAV.

La motivación de este trabajo es que el personal de extinción de incendios pueda usar de manera operativa la información (secuencias de imágenes aéreas), captada por los UAVs de manera que:

- Estas imágenes se emparejen con el ortomapa para poder georreferenciarlas.
- Se combine las imágenes para formar un mosaico, provyendo de una vista cenital de la escena tal y como se encuentra en ese momento.

1.1 Objetivos

El objetivo principal del trabajo es poder georreferenciar imágenes aéreas tomadas por un UAV. Para ello deberemos poner en correspondencia esas imágenes y un ortomapa, ya que de este modo, y gracias a que cada pixel del ortomapa tiene información de su posición, lograremos georreferenciar las imágenes tomadas desde el UAV.

Normalmente las imágenes a emparejar son tomadas desde una misma cámara y se busca emparejarlas de forma secuencial entre ellas para formar un panorama o un mosaico pero en el caso que nos ocupa es distinto por dos razones:

- Las imágenes a emparejar son de cámaras distintas, imágenes aéreas tomadas desde un UAV con una ortofoto generada por el *Institut Cartogràfic de Catalunya*.
- Las tomas de las imágenes corresponden a días distintos y por lo tanto hay diferencias notables entre ellas, tanto en la iluminación en el momento de tomar la imagen como en la vegetación (diferentes estaciones del año) o en la posibilidad de que haya nuevos elemento en ellas.

Debido a esto es necesario hacer una un estudio comparativo e implementar varios detectores-descriptores de características para hallar cuál obtiene un mejor resultado en nuestro caso.

Para lograr abordar el objetivo principal se ha dividido el trabajo en tres subobjetivos:

- Evaluar los detectores-descriptores de características del estado del arte para comprobar cómo se comporta en el emparejamiento de una imagen con un ortomapa.
- Poner en correspondencia frame con ortomapa: Se emparejarán las características de cada frame con las de su zona correspondiente en el ortomapa. Se estimará la homografía [3].
- Validar del sistema: Se evaluarán los resultados obtenidos para calcular el rendimiento del sistema.

Se puede obtener una visión más clara de los objetivos observando el árbol de objetivos en el apéndice A2. *Árbol de objetivos*.

Este artículo está organizado en los siguientes apartados:

- **Estado del arte:** Donde se describen los detectores y descriptores existentes más populares actualmente.
- **Metodología:** Exposición de la metodología seguida en el desarrollo del proyecto.
- **Experimentos:** Este apartado empieza con una breve descripción del dataset utilizado, posteriormente se hace una descripción detallada de los experimentos realizados para la evaluación de detectores y descriptores y se analizan los resultados obtenidos.
- **Conclusiones:** En este apartado se exponen las conclusiones extraídas del proyecto y se comentan las posibles vías de continuación.

2 ESTADO DEL ARTE

Actualmente existen un buen número de detectores-descriptores de características capaces de relacionar pares de imágenes semejantes con variaciones entre ellas como puede ser rotación, zoom, cambio de iluminación, oclusiones, nuevos elementos en la imagen, etc. [4]. En cambio los trabajos encontrados sobre métodos para relacionar una imagen con un mapa se basan en anotación manual. Por ello ha sido necesario estudiar los métodos de emparejamiento imagen-imagen y comprobar si estos son válidos para emparejar una imagen con un ortomapa.

Todos ellos constan de tres etapas para lograr determinar la correspondencia entre dos imágenes (ver Fig. 1).

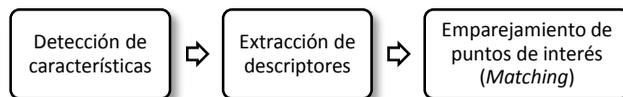


Fig. 1. Proceso de correspondencia de imágenes.

- **Detección de características:** En esta etapa se buscan, mediante un algoritmo de detección, puntos que cumplan unas características concretas dependiendo de cada algoritmo p. ej. bordes, esquinas o regiones. Estos puntos son conocidos como puntos de interés, *keypoints* o *features* indistintamente.
- **Extracción de descriptores:** Durante esta etapa a cada uno de los puntos anteriores se le aplica un algoritmo de descripción el cual genera un vector que lo identifica llamado descriptor local.
- **Emparejamiento de puntos de interés (Matching):** En esta última etapa se comparan los descriptores de los puntos obtenidos en ambas imágenes y se emparejan buscando la distancia mínima entre cada pareja de vectores descriptores.

De entre los detectores-descriptores del estado del arte, se han escogido los más populares para realizar los experimentos del trabajo. En los siguientes apartados se comentan las principales características de cada uno de ellos.

2.1 Detectores de características

2.1.1 BRISK: Binary Robust Invariant Scalable Keypoints

El detector BRISK [5] es capaz de detectar características tipo esquina o *corners* a partir de una imagen de entrada en escala de grises. Esta detección se realiza a múltiples escalas con el fin de lograr invariancia de escala.

2.1.2 FAST: Features from Accelerated Segment Test

El detector FAST [6] es capaz de detectar características tipo esquina o *corners* a partir de una imagen de entrada en escala de grises. Para ello toma la intensidad del píxel a evaluar y la compara con un conjunto de 16 píxeles que lo rodean formando un círculo. Una alternativa para realizar esto más rápido pasa por comparar solo con 4 puntos de esos 16, concretamente los puntos 1, 9, 5 y 13 (ver Fig. 2).

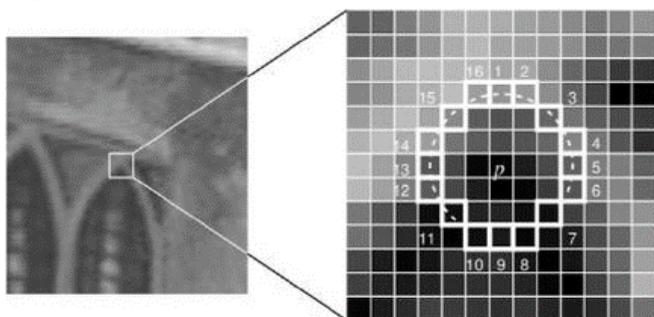


Fig. 2. Imagen que muestra el punto de interés bajo prueba y los 16 píxeles en el círculo (imagen extraída de [6]).

2.1.3 Harris

El detector de esquinas Harris [7] se basa en la función de autocorrelación local de una señal; donde esta mide los cambios locales de la señal desplazando una ventana por una pequeña cantidad de píxeles en diferentes direcciones (ver Fig. 3).

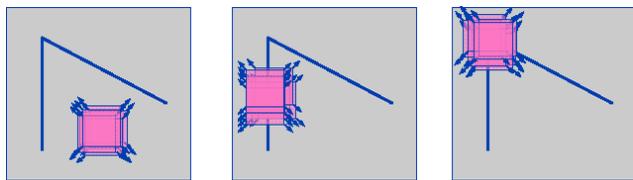


Fig. 3. Imágenes que muestran, de izquierda a derecha: la detección de una zona plana, un contorno y una esquina.

Una esquina o *corner* se caracteriza por ser una región de la imagen con cambios de intensidad en diferentes direcciones.

El método consiste en mirar los cambios de intensidad que se producen en cada píxel para un tamaño determinado de ventana.

En zonas lisas no se producirá ninguna variación en la intensidad; en el caso de un contorno, las variaciones en la intensidad seguirán todas la misma dirección y en el caso de las esquinas se producirán variaciones en la intensidad en múltiples direcciones.

2.1.4 MinEigen: Minimum Eigenvalue

El detector de características MinEigen [8] utiliza el algo-

ritmo de valor propio mínimo desarrollado por J. Shi y C. Tomasi para encontrar puntos de interés o características. Como todos los detectores anteriores, detecta características tipo esquina en imágenes en escala de grises.

2.1.5 MSER: Maximally Stable Extremal Regions

A diferencia de los anteriores, el detector de características MSER [9] en lugar de buscar esquinas busca regiones también conocidas como *blobs*. Esta técnica fue propuesta por Matas *et al.* para encontrar correspondencias entre elementos de la imagen a partir de dos imágenes en escala de grises con diferentes puntos de vista.

2.1.6 SURF: Speeded-Up Robust Features

El detector de características SURF [10] trata de detectar regiones características en imágenes en escala de grises. Para ello hace uso de una aproximación muy básica de la matriz Hessiana [11].

2.2 Descriptores de características

Una práctica común para representar la descripción de un punto de interés o característica local es organizar la información proporcionada por el algoritmo de descripción como los elementos de un vector único, comúnmente se conoce como vector de características.

A continuación se exponen los cuatro descriptores de características locales más usados actualmente.

2.2.1 BRISK: Binary Robust Invariant Scalable Keypoints

El punto clave del descriptor BRISK [5] es que hace uso de un modelo utilizado para el muestreo de la vecindad del punto significativo.

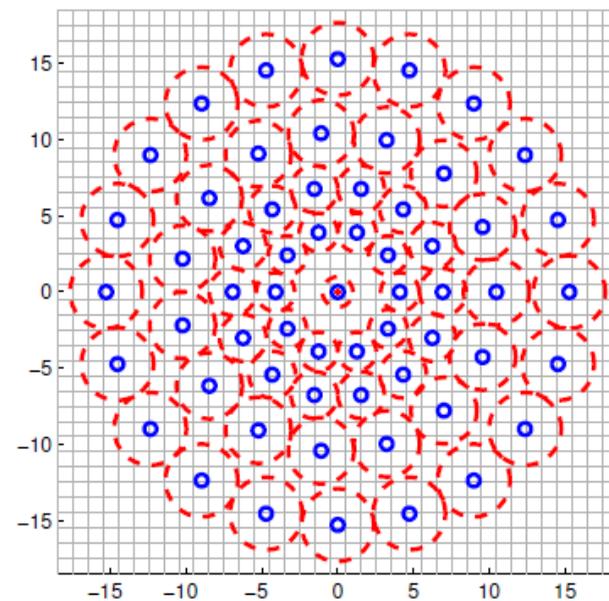


Fig. 4. Patrón de muestreo usado por BRISK para describir un punto de interés (imagen extraída de [5]).

El patrón, que se ilustra en la Fig. 4, define N ubicaciones equidistantes en círculos concéntricos con el punto significativo. Con el fin de evitar los efectos de *aliasing* cuando se muestrea la intensidad de la imagen en un punto p_i en el patrón, se aplica un suavizado gaussiano con desviación estándar y proporcional a la distancia que

haya entre los puntos en el círculo correspondiente.

2.2.2 FREAK: Fast Retina Keypoint

Como el método anterior, el descriptor de características FREAK [12] hace uso de un patrón de píxeles de muestra distribuidos alrededor del punto de interés en forma de círculos concéntricos, no obstante este posee varias diferencias:

- La densidad de píxeles de muestra aumenta conforme nos acercamos al pixel central.
- Las áreas que se tienen en cuenta para el cálculo de la descripción de los píxeles de muestra se solapan unas con otras lo que genera información redundante que mejora el rendimiento del descriptor.
- El área de los píxeles de muestra disminuye de manera exponencial conforme nos vamos acercando al pixel central.

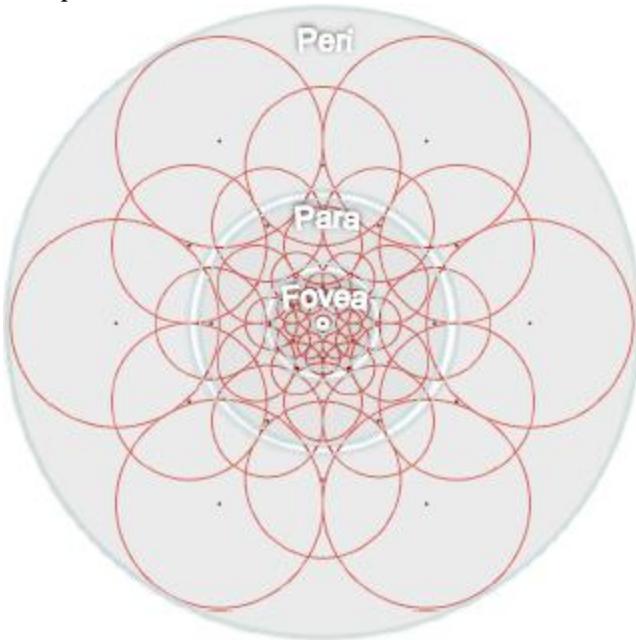


Fig. 5. Patrón de muestreo usado por FREAK para describir un punto de interés (en rojo), sobrepuesta en un esquema de las áreas de la retina humana (imagen extraída de [12]).

En la Fig. 5 se muestra un esquema del patrón de puntos de muestreo empleado por FREAK. Cada círculo representa la desviación estándar del *kernel* de Gauss usado para suavizar los puntos de muestreo correspondientes y de este modo eliminar el ruido.

2.2.3 SURF: Speeded-Up Robust Features

El descriptor SURF [10], describe la distribución del contenido de intensidad dentro un área entorno al punto de interés de forma similar a la información de gradiente extraída por SIFT [13], que es un descriptor que ofrece una buena *performance* pero no puede usarse de forma libre ya que está patentado. En este caso, SURF a diferencia de SIFT, se basa en la distribución del *wavelet* de Haar de primer orden en dirección X e Y en lugar de gradiente.

Por otro lado utiliza imágenes integrales para mejorar la velocidad y usa solo 64 dimensiones. Esto reduce el tiempo para el cálculo de características y mejora simul-

táneamente la robustez.

El descriptor emplea dos pasos para lograr describir un punto de interés:

- Asignación de la orientación para lograr invariancia a rotaciones.
- Cálculo del descriptor basándose en la suma de las respuestas al *wavelet* de Haar (ver Fig. 6).

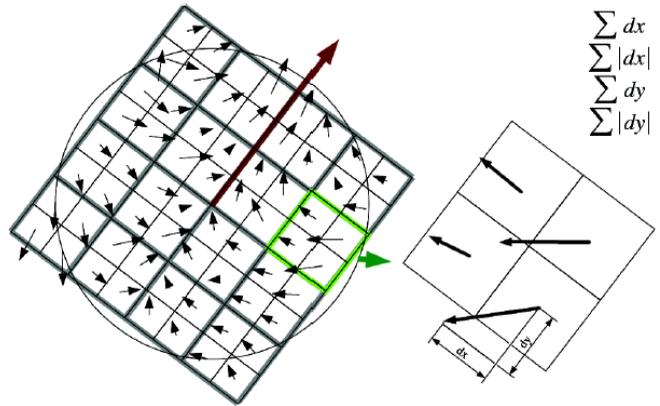


Fig. 6. Rejilla para el cálculo del descriptor SURF, cada flecha representa la respuesta al *wavelet* de Haar para esa región (imagen extraída de [10]).

2.2.4 Block

El método Block simplemente extrae los valores de intensidad de los píxeles en bruto una región cuadrada alrededor de cada punto de interés. Este es el descriptor más trivial. Es el más rápido de calcular, pero es mucho menos robusto que los otros en términos de precisión a la hora de buscar los emparejamientos.

3 METODOLOGÍA

El trabajo se ha organizado en función de los objetivos planteados, dividiéndolo en varias fases. Se ha seguido un desarrollo incremental e iterativo en el cual se han analizado críticamente cada uno de los resultados obtenidos.

Fase 1: Investigación de documentación relacionada

- En esta fase se ha buscado y estudiado información sobre los detectores - descriptores del estado del arte.

Fase 2: Preparación del entorno de trabajo

- En esta fase se ha determinado tanto la tecnología cómo las herramientas que se han usado.
- También se ha procedido a la instalación del software necesario para la codificación de los experimentos. El IDE elegido ha sido Matlab, ya que es uno de los más potentes actualmente para trabajar con imágenes y su librería tiene implementados los algoritmos más populares de detección y descripción de puntos de interés.

Fase 3: Evaluación de los detectores-descriptores

- Se ha determinado el *dataset* a usar el cual está formado por un conjunto de imágenes tomadas por un UAV y un ortomapa del *Institut Cartogràfic de Catalunya*.

- Se ha anotado el emparejamiento ideal entre frame y ortomapa seleccionando puntos característicos en la captura y el ortomapa para determinar qué región del ortomapa corresponde con cada captura.
- Se han evaluado los detectores de características analizando cual consigue detectar un mayor número de características en el mismo sitio en un frame y en su zona correspondiente del ortomapa.
- Se han evaluado los descriptores de características calculándolos en las coordenadas correspondientes entre frame y ortomapa y analizando sus semejanzas.

4 EXPERIMENTOS REALIZADOS

En este apartado se expondrá el trabajo realizado durante la fase 3 del proyecto, "Evaluación de los detectores-descriptores" y los resultados que se han obtenido. Los experimentos se dividen en dos apartados: Evaluación de detectores y evaluación de descriptores.

4.1 Dataset

El conjunto de imágenes seleccionado para los experimentos está formado por 11 capturas extraídas de una secuencia de vídeo. Esta secuencia de vídeo se realizó haciendo uso de un UAV (*Unmanned Aerial Vehicle*) durante un vuelo realizado sobre el campo de vuelo de Sant Cugat de Vallès y por un ortomapa de esa zona obtenido del *Institut Cartogràfic de Catalunya* con una resolución de 15 cm/píxel (ver apéndice A3. *Imágenes del dataset*).

Las imágenes originales presentaban un efecto conocido como efecto peine [14]. Esto es debido a que la captura del vídeo se ha hecho en movimiento y esto junto con el método de doble barrido que utiliza la cámara para capturar los *frames* suele provocar el efecto que se puede observar en la Fig. 7 sobre una de las capturas. En la Fig. 8 se muestra cómo queda después de editarla extrayendo de ella las filas impares y haciendo lo mismo con las columnas impares para mantener la proporción original de la cámara expensas de reducir la resolución.



Fig. 7. Ejemplo de captura con efecto peine.



Fig. 8. Captura muestreada corrigiendo el efecto peine a expensas de reducir la resolución.

Una vez corregido el efecto peine de todas las imágenes del dataset, se ha generado un emparejamiento de puntos ideal entre las capturas y el ortomapa. Seleccionando manualmente 6 puntos en la captura y seis puntos en el ortomapa haciendo coincidir en ambos casos esquinas o puntos reconocibles para, de este modo, poder calcular la homografía de cada captura; es decir la matriz de transformación que aplicada a las coordenadas de un punto de una captura te devuelve las coordenadas de ese punto en el ortomapa.

Con estas matrices de transformación se ha creado un *script* para formar un mapa a partir de las homografías de las capturas. Este *script* se ha creado modificando el ejemplo disponible en Matlab R2014b "Feature Based Panoramic Image Stitching" [15].

Originalmente este ejemplo fue creado para unir varias fotografías tomadas de forma solapada y en orden, para formar un panorama, uniendo todas ellas y haciéndolas encajar. En este caso lo que se desea crear es algo diferente ya que todas las imágenes se deben corresponder con el ortomapa y no necesariamente con la anterior imagen, por lo que este ejemplo se ha modificado para adaptarlo a las necesidades del proyecto, de tal manera que ahora la imagen del ortomapa se considera como la primera imagen y en cada iteración se va correspondiendo una captura con el ortomapa.

La Fig. 9 muestra el primer resultado obtenido. Claramente el encaje entre capturas es pobre, lo que refleja una anotación de puntos entre capturas y ortomapa poco precisa. Para solucionar este problema se repitió la anotación manual de puntos, esta vez haciendo *zoom* en la captura y el ortomapa para realizar una anotación a nivel subpíxel.

El resultado después de la segunda anotación se puede ver en la Fig. 10, en la que se observa una mejora considerable con respecto al primer resultado.

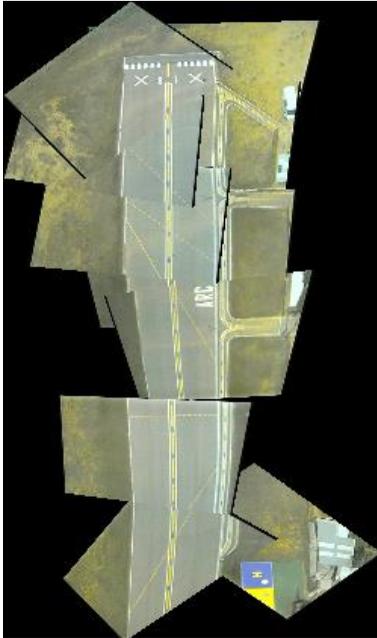


Fig. 9. Primer mapa, generado a partir de la homografía de las capturas.



Fig. 10. Segundo mapa, generado tras la anotación a nivel subpixel.

4.2 Experimentos

Para emparejar capturas y ortomapa se parte de la hipótesis de que una captura y su zona correspondiente en el ortomapa son bastante similares. En consecuencia, deberían generar unos puntos de interés en las mismas zonas o en zonas cercanas. Por otro lado, la descripción de la región centrada en cada uno de esos puntos debería ser similar.

Como en este caso las capturas no han sido tomadas

con la misma cámara ni en el mismo momento que el ortomapa, se plantean los siguientes dos experimentos para determinar que detector y que descriptor se adapta mejor a esta situación y logra obtener mejores resultados.

4.2.1 Evaluación de detectores

Se han seleccionado los detectores BRISK, FAST, Harris, MinEigen, MSER y SURF por ser estos los más populares del estado del arte.

El experimento ideado para evaluar los detectores constaba inicialmente de tres etapas, a continuación se detalla los pasos del proceso para una de las capturas del dataset:

1. Utilizando la homografía ideal calculada en tareas anteriores, se obtiene la localización de la región correspondiente del ortomapa. Estas delimitan la zona del ortomapa que corresponde a la captura, en la que se deben buscar los puntos de interés.
2. Se ejecuta el detector sobre la captura y sobre la zona del ortomapa correspondiente, indicando las coordenadas obtenidas en el anterior paso mediante el parámetro ROI [16], para obtener los puntos de interés en ambas.
3. Se calcula la distancia euclídea entre cada punto de la captura y del ortomapa y se cuenta cuantas parejas de puntos hay que estén a cierta distancia (p.ej. cuantas parejas de puntos se han encontrado que estén a 2 píxeles de distancia entre ellos). Para el experimento se han contabilizado parejas distanciadas entre 0 y 4 píxeles ya que se ha considerado que a más de 4 píxeles no se trataba del mismo punto de interés.

Con este primer experimento se observaban valores por debajo de 25 % de coincidencias entre captura y ortomapa para radio 4. Esto ocurría de manera indistinta al detector de características que se emplease, por lo que el problema se supuso que lo originaban las imágenes.

Tras observar varias capturas junto al recorte del ortomapa correspondiente, se llegó a la conclusión de que el problema venía de la diferencia de resolución entre captura y ortomapa. Esto se debe a que las imágenes se tomaron a distintas alturas.

En la Fig. 11 se puede ver un ejemplo de la diferencia de resolución entre captura y ortomapa en la que se aprecia como el ortomapa presenta un pixelado más visible. Hay que tener en cuenta la imagen del recorte del ortomapa (derecha) está aumentada, en este documento, para verla del mismo tamaño que la captura.

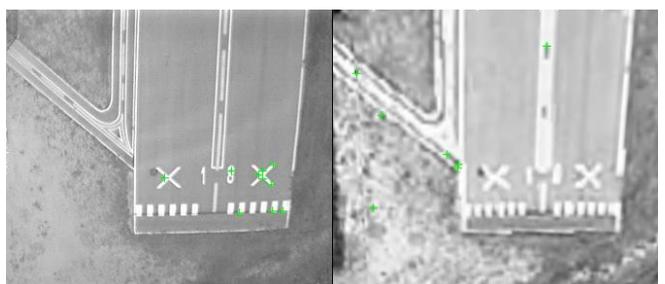


Fig. 11. Ejemplo de diferencia de resolución entre captura y ortomapa.

Por lo tanto se decidió aplicar un filtro de suavizado a ambas imágenes y redimensionar la captura para igualar la escala a la del ortomapa. De modo que el experimento para evaluar los detectores finalmente consta de cinco etapas, a continuación se detalla los pasos añadidos al proceso para una de las capturas del dataset:

- Esta etapa se realiza antes de la etapa 1. Se aplica un filtro de suavizado a la captura y otro al ortomapa.
- Esta etapa se realiza entre la etapa 1 y la 2. Se calcula el factor de escala el ortomapa y la captura usando la matriz de homografía para calcular la distancia entre dos puntos de la captura y esos mismos puntos trasladados al ortomapa y a continuación se le aplica a esta captura la función `imresize` [17] con ese factor de escala para equiparar la resolución de esta a la del ortomapa.

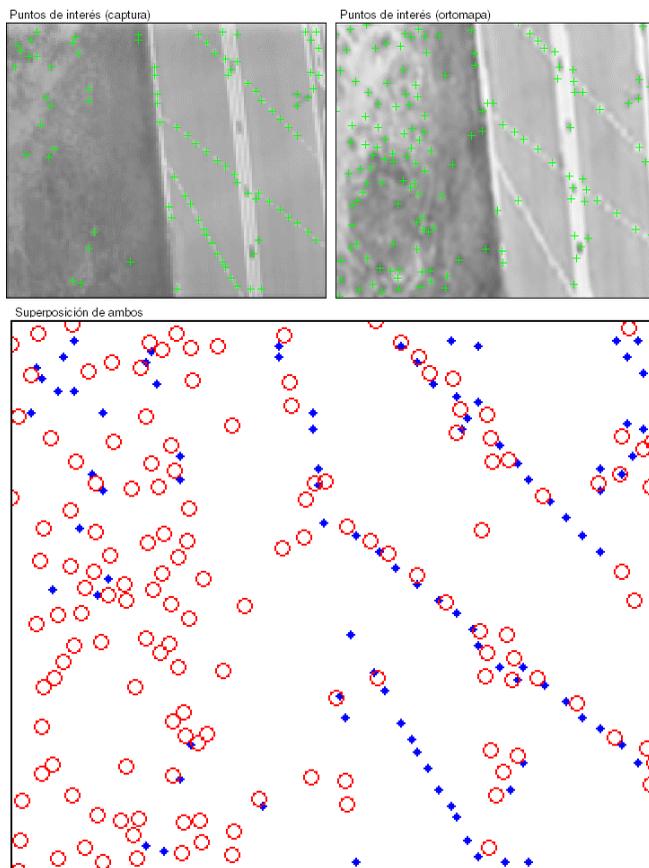


Fig. 12. Ejemplo de puntos encontrados con el detector FAST (captura: `MyVideo_14_58_58_219_062.png`).

En la Fig. 12 se muestra un ejemplo de la ejecución del detector FAST sobre una captura y su zona correspondiente en el ortomapa. En la primera imagen se ven los puntos de interés detectados en la captura, en la segunda los detectados en el ortomapa y en la tercera imagen una superposición de ambos donde los puntos son los detectados en la captura y los círculos los detectados en el ortomapa. Se observa que en la imagen del ortomapa se han detectado muchos más puntos de interés, sobre todo en la zona fuera de la pista de aterrizaje que es menos homogénea ya que contiene trozos sin arbustos que en la captura no hay. No obstante en el resto de la imagen del orto-

mapa se observa también un contraste importante entre píxeles vecinos, haciéndola abrupta y poco natural.

Finalmente, para poder analizar los resultados, estos se han colocado en unas tablas de Excel, una por detector evaluado. En el apéndice A4. *Ejemplo de la tabla de resultados de evaluar los detectores (Harris)* se muestra la forma de la tabla, la cual contiene los siguientes datos:

- Detector: nombre del detector evaluado.
- Captura: nombre de la captura del dataset.
- # Puntos captura: número de puntos encontrados en la captura.
- # Puntos ortomapa: número de puntos encontrados en el ortomapa.
- Matches R0 – R4: número de puntos encontrados en ambas imágenes desde R0 (en el mismo píxel) hasta R4 (a cuatro píxeles de distancia).
- Porcentaje R0 – R4: Porcentaje que representan las cinco columnas anteriores sobre el número de puntos encontrados en la captura u ortomapa (seleccionando el que menos puntos contenga).
- En mismas columnas anteriores, los valores de la última fila muestran la media de estos porcentajes para cada radio de observación.
- En el área inferior se muestra tanto para la captura como para el ortomapa, primero el valor de configuración del filtro de suavizado (`sigCapture` y `sigOrtomapa`) y a continuación los valores de los parámetros específicos de cada detector.

Para analizar los resultados obtenidos del experimento de evaluación de los detectores se ha cogido la última fila de la tabla de resultados de cada descriptor (ver apéndice A4. *Ejemplo de la tabla de resultados de evaluar los detectores (Harris)*), la cual muestra el promedio de porcentajes de coincidencias de todas las capturas del dataset para los diferentes radios de observación, y se han incluido en la tabla resumen que muestra la Fig. 13. Esta tabla se ha coloreado cada columna en una escala de grises para diferenciar a simple vista que detector de características obtiene mejor resultado para cada radio de observación. Concretamente cuanto más oscura sea la casilla mejor resultado refleja.

Detector	Radio 0	Radio 1	Radio 2	Radio 3	Radio 4
BRISK	0,5185	1,4338	5,0494	9,4929	13,2290
FAST	1,0533	6,5355	15,0774	20,8546	31,0543
Harris	3,1623	11,5704	19,1091	27,6552	36,1912
MinEigen	3,5459	9,2480	18,5851	28,6282	35,5408
MSER	0,5682	1,2160	4,2976	7,0641	13,5384
SURF	1,4242	4,0992	9,4229	14,2024	19,0637

Fig. 13. Tabla resumen de los resultados del experimento para los detectores de características.

Como se puede observar los detectores Harris y MinEigen destacan por encima del resto. En ambos casos se trata de detectores de esquinas, se deduce por tanto que en el caso que se estudia en el proyecto, imágenes tomadas desde distintas cámaras y a distinta distancia del suelo, son más efectivos los detectores de esquinas que los de regiones. Esto principalmente se debe a la diferencia

de contraste entre píxeles vecinos que hace que no se encuentren zonas homogéneas donde localizar regiones.

Aun así tanto Harris como MinEigen no alcanzan el 40% de coincidencias, lo cual es un resultado bastante bajo en términos absolutos.

4.2.2 Evaluación de descriptores

Se han seleccionado los descriptores BRISK, FREAK, SURF y Block por ser esos los más populares del estado del arte.

El experimento ideado para evaluar los descriptores constaba inicialmente de cinco etapas, a continuación se detalla los pasos del proceso para una de las capturas del dataset:

1. Se aplica un filtro de suavizado a la captura y otro al ortomapa.
2. Se calcula el factor de escala del ortomapa y la captura usando la matriz de homografía para calcular la distancia entre dos puntos de la captura y esos mismos puntos trasladados al ortomapa y a continuación se le aplica a esta captura la función *imresize* [17] con ese factor de escala para equiparar la dimensión de esta a la del ortomapa.
3. Se determinan los puntos que en los que se calculará el descriptor. Inicialmente establecieron los anotados para el emparejamiento ideal entre captura y ortomapa. Y se generan mediante el algoritmo a evaluar un vector descriptor por cada punto.
4. A partir de las descripciones de los puntos se calcula la diferencia entre los puntos de la captura y los del ortomapa usando la suma de las diferencias al cuadrado (SSD) [18] para obtener una matriz cuadrada de distancias entre los vectores descriptores de los píxeles de la captura y los vectores descriptores de los píxeles del ortomapa. Se ha utilizado SSD ya que esta métrica es la que se utiliza habitualmente para comparar vectores descriptores.
5. Por último se busca el valor más bajo de cada fila, es decir que pixel del ortomapa tiene menor distancia (se parece más) a cada pixel de la captura, posteriormente se cuentan cuántos de estos valores se encuentran en la diagonal, lo que quiere decir que son entre el mismo pixel de la captura que del ortomapa. Para tener una forma objetiva e independiente del número de puntos usados de evaluar los descriptores se ha calculado el tanto por ciento de puntos en la diagonal para cada imagen del dataset y para cada descriptor.

Al ejecutar este experimento se encontró que para algunas imágenes del dataset no se retornaba la descripción de los 6 puntos del emparejamiento ideal. Esto se debía a que los descriptores utilizan para describir un punto de interés un área que rodea a ese pixel.

En el caso de BRISK y FREAK se tiene en cuenta una circunferencia de 12 y 18 píxeles de radio respectivamente centrada en el punto de interés y en el caso de SURF y Block se emplean los píxeles que quedan dentro de un cuadrado de lado 8 en el primer caso y 11 en el segundo. Por lo que al intentar describir un punto de interés en el

que esa área se saliera de la imagen, este se trataba como punto no válido y no se describía.

Para subsanar este problema se decidió no hacer uso de los puntos del emparejamiento ideal para comparar los descriptores, sino de una rejilla de puntos que se adaptara a cada captura en función del tamaño de esta y asegurara que el cálculo del descriptor fuera posible para todos los puntos (ver Fig. 14).

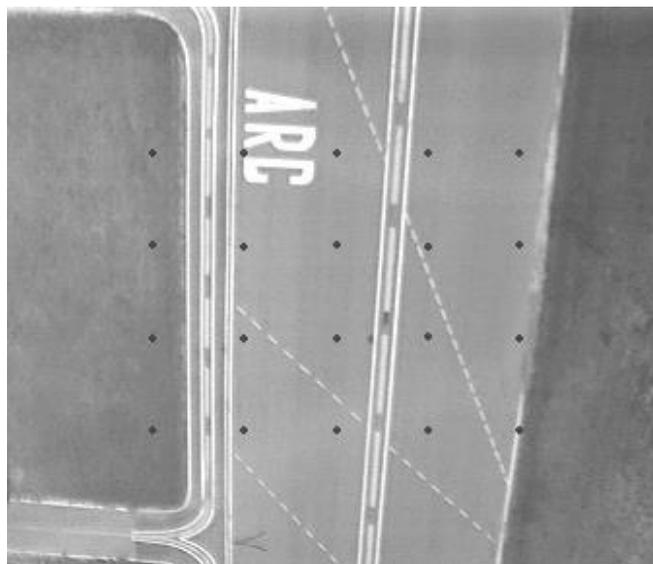


Fig. 14. Ejemplo de rejilla de puntos generada para el cálculo de los descriptores.

Por lo tanto se ha modificado la etapa 3 del procedimiento para evaluar los descriptores quedando de la siguiente forma:

- Se crea una rejilla de coordenadas para determinar qué puntos se van a describir para cada captura, esta rejilla de coordenadas se calcula sobre la captura ya que esta es más pequeña que el ortomapa y así se logra que ningún punto quede fuera de la zona válida de puntos. Se describen los puntos que corresponden a estas coordenadas tanto para la captura como para la zona del ortomapa correspondiente. Obteniendo un vector descriptor por cada punto.

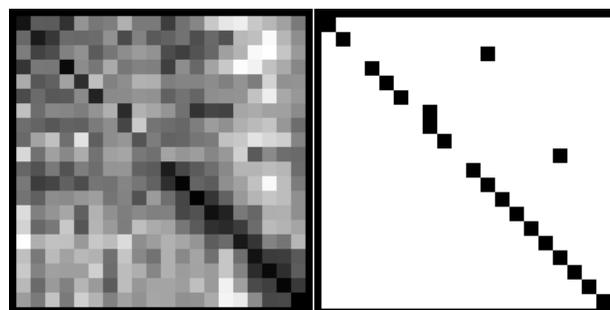


Fig. 15. Ejemplo de matriz de distancias entre píxeles (izquierda) y la selección de las distancias más cortas (derecha).

En la Fig. 15 se muestra un ejemplo donde se puede ver, en la imagen izquierda, la matriz de distancias entre los puntos encontrados en la captura (filas) y los encontrados en la captura (columnas).

trados en el ortomapa (columnas). En la imagen derecha se muestran marcadas solo las distancias más cortas. Claramente se identifican 3 casos en los que el pixel más cercano al pixel de la captura no es su equivalente en el ortomapa.

Para analizar los resultados del experimento de evaluación de descriptores se ha calculado el porcentaje de distancias mínimas en la diagonal de cada imagen con cada detector (ver Fig. 16)

Captura / Descriptor	BRISK	FREAK	SURF	Block
MyVideo_14_57_46_342_061	26,67%	26,67%	46,67%	6,67%
MyVideo_14_58_06_234_061	60,00%	30,00%	40,00%	5,00%
MyVideo_14_58_08_670_063	35,00%	10,00%	45,00%	10,00%
MyVideo_14_58_17_391_063	65,00%	30,00%	20,00%	10,00%
MyVideo_14_58_21_172_062	50,00%	15,00%	45,00%	5,00%
MyVideo_14_58_29_389_062	35,00%	35,00%	40,00%	0,00%
MyVideo_14_58_37_096_064	50,00%	35,00%	50,00%	5,00%
MyVideo_14_58_45_420_-938	40,00%	30,00%	40,00%	15,00%
MyVideo_14_58_58_219_062	80,00%	65,00%	50,00%	10,00%
MyVideo_14_59_02_109_061	85,00%	50,00%	55,00%	15,00%
MyVideo_14_59_06_686_061	55,00%	40,00%	30,00%	0,00%
Porcentaje de distancias mínimas en la diagonal	52,88%	33,33%	41,97%	7,42%

Fig. 16. Tabla resumen de los resultados del experimento para los descriptores de puntos de interés.

Se puede ver claramente como el descriptor BRISK es superior al resto de descriptores para la mayoría de capturas del dataset y por lo tanto es también superior en la media calculada de todas las imágenes. Este valor muestra que para más del 50% de las descripciones de los puntos de la capturas la descripción que menos difiere es la del punto análogo en el ortomapa.

Con esto se determina que BRISK es el mejor descriptor para el caso que nos ocupa, imágenes tomadas con distinta cámara, a distinta altura y en distintos días.

5 CONCLUSIONES

A lo largo de este proyecto se han realizado varias tareas que han llevado a obtener los siguientes resultados:

- **Análisis del estado del arte:** Se han analizado los detectores-descriptores del estado del arte y se han seleccionado los más populares para realizar los experimentos
- **Generación de dataset:** Se han seleccionado 11 capturas obtenidas de un video tomado por un UAV, con las cuales iban a realizar las pruebas para resolver los objetivos planteados. Se ha generado un emparejamiento ideal para obtener la matriz de homografía entre cada captura y la zona correspondiente del ortomapa.
- **Evaluación de los detectores:** Se han evaluado los detectores BRISK, FAST, Harris, MinEigen, MSER y SURF para determinar cuál consigue detectar un mayor número de características en el mismo sitio en

un frame y en su zona correspondiente del ortomapa. Tras lo cual se ha llegado a la conclusión de que para un problema de las características del que se ha tratado en este proyecto, el detector de características más adecuado es Harris el cual supera mínimamente a MinEigen en la similitud a la hora de encontrar los mismos puntos característicos entre una captura y el ortomapa.

- **Evaluación de los descriptores:** Se han evaluado los descriptores BRISK, FREAK, SURF y Block para determinar cuál de ellos obtiene vectores de descripción más parecidos entre los puntos seleccionados en las capturas y el ortomapa. Se ha llegado a la conclusión que para el caso tratado en el proyecto, el descriptor de puntos de interés más adecuado es BRISK con un 52,88% de media de aciertos en el emparejamiento de vectores de descripción.

En resumen, tras los experimentos realizados, se puede afirmar, aunque obtengan un resultado discreto en términos absolutos, que la mejor pareja de detector-descriptor del estado del arte para emparejar imágenes aéreas tomadas con distintas cámaras, a distintas alturas y en distintos días, es la formada por el detector Harris y el descriptor BRISK.

5.1 Vías de continuación

Por falta de tiempo, debido a que el estudio de detectores y descriptores ha supuesto más trabajo del previsto, no se han podido completar dos de los subobjetivos inicialmente planteados en las siguientes fases:

- **Fase 4:** Puesta en correspondencia entre capturas y ortomapa.
- **Fase 5:** Validación del sistema.

Por lo tanto una posible vía de continuación podría ser completar estas 2 fases.

Para cumplir la primera se podría reutilizar el código de la tarea "Generación de emparejamiento ideal frame-ortomapa" de la fase 3 para hacer corresponder las capturas con el ortomapa y generar un mapa a partir de ellas, haciendo uso del detector Harris y descriptor BRISK, ya que estos han sido los que han demostrado ir mejor.

Para cumplir la segunda fase se debería haber analizado el resultado del anterior para determinar si el rendimiento del sistema satisface o no las expectativas generadas al principio del proyecto.

Como idea interesante a añadir se podría repetir los experimentos con diferentes datasets p. ej. zonas urbanas, agrícolas... para determinar si los resultados del estudio son los mismos en zonas naturales y en zonas que contienen objetos construidos por el hombre.

Otra posible vía de continuación podría ser intentar mejorar el resultado de la evaluación de descriptores, seleccionando de la captura solamente los puntos que no generen ambigüedad entre ellos, es decir los que sus vectores descriptores disten más entre sí. Con esto posiblemente se consiga generar menos emparejamientos erróneos con los puntos del ortomapa.

AGRADECIMIENTOS

Al mi tutor Daniel Ponsa por la ayuda semana tras semana en el desarrollo de este proyecto así como a los consejos que me ha dado para mejorar los informes y el artículo.

A Felipe Lumbreras por ayudar todo lo posible en los momentos en los que encontraba un error y no sabía el origen.

A mi familia y amigos por entender el poco tiempo que he podido dedicarles durante estos meses y que pienso recuperar en los siguientes.

BIBLIOGRAFÍA

- [1] Fireguard Project (Lab-UAS). Technical Innovation Cluster in Aeronautical Management, 2013 [consulta 23-01-2015]. Disp. en: <http://centresderecerca.uab.cat/aeronauticalcluster/content/fireguard-project-lab-uas>
- [2] FireWATCHER, Fire Warning by Aerial Terrain Control of Hot Embers Regions. Centre de Visió per Computador - Advanced Driver Assistance Systems Group, 2015 [consulta 23-01-2015]. Disp. en: <http://www.cvc.uab.es/adas/projects/firewatcher/>
- [3] E. Dubrofsky, "Homography Estimation." The university of British Columbia (Vancouver), 2009.
- [4] T. Tuytelaars and K. Mikolajczyk, "Local Invariant Feature Detectors: A Survey." Foundations and Trends in Computer Graphics and Vision Vol. 3, No. 3 (2007) 177–280.
- [5] S. Leutenegger, M. Chli and R. Y. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints." Autonomous Systems Lab, ETH Zürich, 2011.
- [6] FAST Algorithm for Corner Detection. OpenCV 3.0.0-dev documentation, OpenCV, 2011-14 [consulta 25-01-2015]. Disp. en: http://docs.opencv.org/trunk/doc/py_tutorials/py_feature2d/py_fast/py_fast.html
- [7] C. Harris and M. Stephens, "A combined corner and edge detector." Proceedings of the 4th Alvey Vision Conference, August 1988, pp. 147-151.
- [8] J. Shi and C. Tomasi, "Good Features to Track." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 1994, pp. 593–600.
- [9] J. Matas, O. Chum, M. Urban, and T. Pajdla. "Robust wide baseline stereo from maximally stable extremal regions." Proceedings of British Machine Vision Conference, pages 384-396, 2002.
- [10] Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool. "SURF: Speeded Up Robust Features." Computer Vision and Image Understanding (CVIU). Vol. 110, No. 3, pp. 346–359, 2008.
- [11] Matriz Hessiana. Wikipedia, 2014 [consulta 26-01-2015]. Disp. en: http://es.wikipedia.org/wiki/Matriz_hessiana
- [12] A. Alahi, R. Ortiz and P. Vandergheynst. "FREAK: Fast Retina Keypoint." Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, 2012.
- [13] D.G. Lowe, "Object recognition from local scale-invariant features," Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol.2, no., pp.1150,1157 vol.2, 1999 doi: 10.1109/ICCV.1999.790410
- [14] Efecto peine. Wikipedia, 2013 [consulta 28-01-2015]. Disp. en: http://es.wikipedia.org/wiki/Efecto_peine
- [15] Feature Based Panoramic Image Stitching. The MathWorks Inc., R2014b Documentation, 2014. [consulta 15-11-2014]. Disp. en: <http://es.mathworks.com/help/vision/examples/feature-based-panoramic-image-stitching.html>
- [16] Specifying a Region of Interest (ROI). The MathWorks Inc., R2014b Documentation, 2014. [consulta 05-12-2014]. Disp. en: <http://es.mathworks.com/help/images/specifying-a-region-of-interest-roi.html>
- [17] Reescalado de una imagen. Imresize. The MathWorks Inc., R2014b Documentation, 2014. [consulta 05-12-2014]. Disp. en: <http://es.mathworks.com/help/images/ref/imresize.html>
- [18] Sum of Squared Differences (SSD): Correlation based similarity measures-Summary. Siddhant ahuja (sid), Wordpress.com, Abril 2010. [consulta 23-12-2014]. Disp. en: <https://siddhantahuja.wordpress.com/tag/sum-of-squared-differences/>

APÉNDICES

A1. Hexarotor y sistema de adquisición

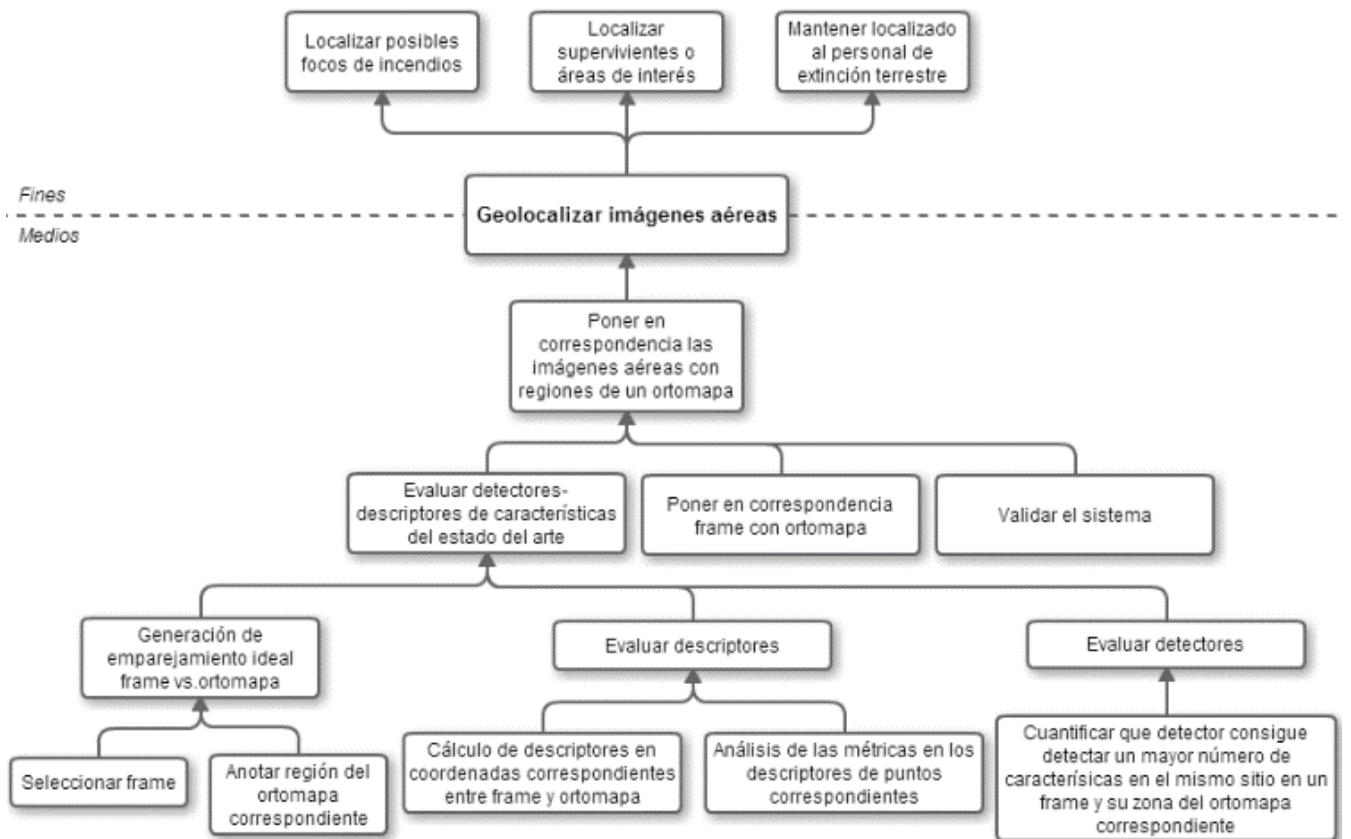


Hexarotor utilizado para el vuelo.

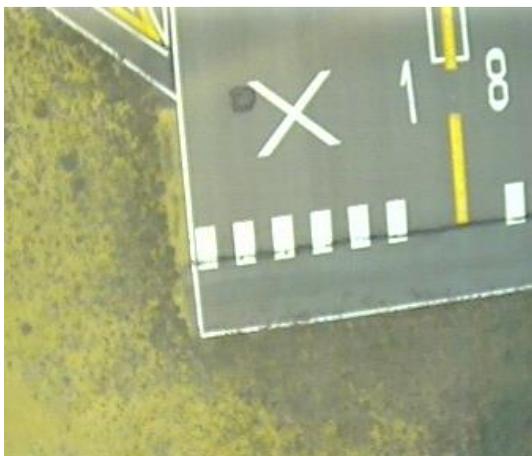


Sistema de adquisición de imágenes.

A2. Árbol de objetivos



A3. Imágenes del dataset



MyVideo_14_57_46_342_061.png



MyVideo_14_58_17_391_063.png



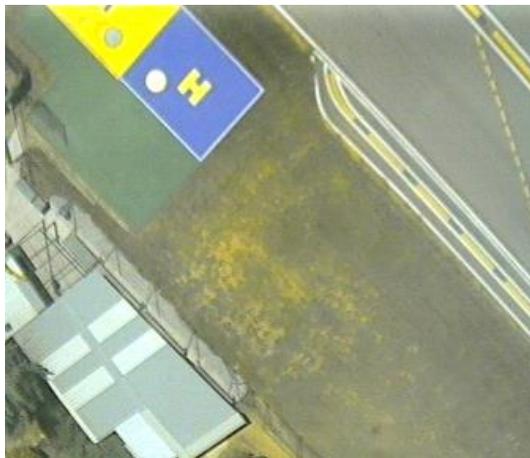
MyVideo_14_58_06_234_061.png



MyVideo_14_58_21_172_062.png



MyVideo_14_58_08_670_063.png



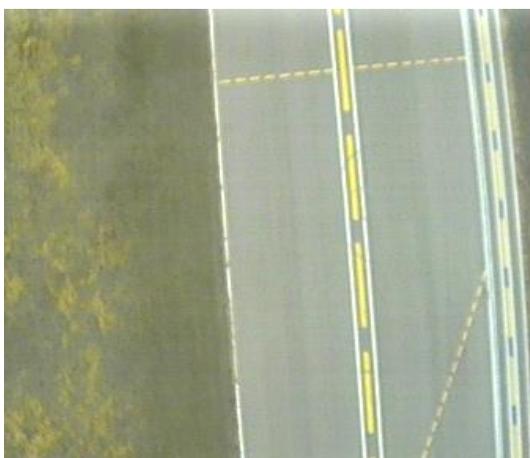
MyVideo_14_58_29_389_062.png



MyVideo_14_58_37_096_064.png



MyVideo_14_59_02_109_061.png



MyVideo_14_58_45_420_-938.png



MyVideo_14_59_06_686_061.png



MyVideo_14_58_58_219_062.png



CampVolSantCugatWGS84.tif

A4. Ejemplo de la tabla de resultados de evaluar los detectores (Harris)

Detector	Captura	# Puntos captura	# Puntos ortomapa	Matches R0	Matches R1	Matches R2	Matches R3	Matches R4	Porcentaje R0	Porcentaje R1	Porcentaje R2	Porcentaje R3	Porcentaje R4
Harris	MyVideo_14_57_46_342_061	39	56	0	3	4	7	9	0,0000	7,6923	10,2564	17,9487	23,0769
Harris	MyVideo_14_58_06_234_061	64	133	4	16	17	25	30	6,2500	25,0000	26,5625	39,0625	46,8750
Harris	MyVideo_14_58_08_670_063	49	55	0	1	2	8	10	0,0000	2,0408	4,0816	16,3265	20,4082
Harris	MyVideo_14_58_17_391_063	58	37	1	5	9	10	14	2,7027	13,5135	24,3243	27,0270	37,8378
Harris	MyVideo_14_58_21_172_062	46	24	0	0	6	9	11	0,0000	0,0000	25,0000	37,5000	45,8333
Harris	MyVideo_14_58_29_389_062	71	75	3	5	6	9	15	4,2254	7,0423	8,4507	12,6761	21,1268
Harris	MyVideo_14_58_37_096_064	31	54	0	2	4	5	8	0,0000	6,4516	12,9032	16,1290	25,8065
Harris	MyVideo_14_58_45_420_-938	88	99	5	9	15	18	28	5,6818	10,2273	17,0455	20,4545	31,8182
Harris	MyVideo_14_58_58_219_062	83	88	1	13	17	21	26	1,2048	15,6627	20,4819	25,3012	31,3253
Harris	MyVideo_14_59_02_109_061	164	131	4	17	32	46	62	3,0534	12,9771	24,4275	35,1145	47,3282
Harris	MyVideo_14_59_06_686_061	60	138	7	16	22	34	40	11,6667	26,6667	36,6667	56,6667	66,6667
Promedio									3,1623	11,5704	19,1091	27,6552	36,1912
Captura: sigCapture = No se aplica; 'FilterSize',3; 'MinQuality',0.01													
Ortomapa: sigOrto = No se aplica; 'FilterSize',3; 'MinQuality',0.01													